On false discovery rate thresholding for classification under sparsity

Pierre Neuvial, Etienne Roquain

July 1, 2011

Abstract

We study the properties of false discovery rate (FDR) thresholding, viewed as a classification procedure. The "0"-class (null) is assumed to have a known, symmetric log-concave density while the "1"-class (alternative) is obtained from the "0"-class either by translation (location model) or by scaling (scale model). Furthermore, the "1"-class is assumed to have a small number of elements w.r.t. the "0"-class (sparsity). Non-asymptotic oracle inequalities are derived for the excess risk of FDR thresholding. In a regime where Bayes power is away from 0 and 1, these inequalities lead to explicit rates of convergence of the excess risk to zero. Moreover, these theoretical investigations suggest an explicit choice for the nominal level α_m of FDR thresholding, in function of m. Our oracle inequalities show theoretically that the resulting FDR thresholding adapts to the unknown sparsity regime contained in the data. This property is illustrated with numerical experiments, which show that the proposed choice of α_m is relevant for a practical use.

1 Introduction

1.1 Background

The false discovery rate (FDR) has become a standard for analyzing many types of data, such as microarray or neuro-imaging. Albeit motivated by pure testing considerations, recent studies have shown that the Benjamini Hochberg FDR controlling procedure proposed by [2] enjoys remarkable properties as a detection procedure [9] and as an estimation procedure [1, 10]. More specifically, it turns out to be adaptive to the amount of "signal" contained in the data, which has been referred to as "adaptation to unknown sparsity".

Recently, an important theoretical breakthrough has been made with the study of FDR thresholding in a classification framework, where asymptotic results were proved in a Gaussian scale model [6] (see also [17] and [7]). The present paper extends this work by studying the adaptation to unknown sparsity of FDR thresholding non-asymptotically and in more general models (location/scale models with symmetric log-concave densities, see Section 1.6 for a detailed comparison to [6]).

1.2 Initial setting

Let $(X_i, H_i) \in \mathbb{R} \times \{0, 1\}$, $1 \leq i \leq m$, be m i.i.d. variables. Assume that the sample $X_1, ..., X_m$ is observed without the labels $H_1, ..., H_m$ and that the distribution of X_1 conditionally on $H_1 = 0$ is known a priori. We consider the following general classification problem: build a

(measurable) classification rule $\hat{h}_m : \mathbb{R} \to \{0,1\}$, depending on $X_1, ..., X_m$, such that, for a new labeled data point $(X_{m+1}, H_{m+1}) \sim (X_1, H_1)$ independent of $(X_i, H_i)_{1 \leq i \leq m}$, the (integrated) misclassification risk

$$R_m(\hat{h}_m) = \mathbb{P}(\hat{h}_m(X_{m+1}) \neq H_{m+1}) \tag{1}$$

is as small as possible.

The distribution of (X_1, H_1) is assumed to belong to a specific parametric subset of distributions on $\mathbb{R} \times \{0,1\}$, which is defined as follows:

(i) the distribution of H_1 is such that the (unknown) mixture parameter $\tau_m = \pi_{0,m}/\pi_{1,m}$ satisfies $\tau_m > 1$, where $\pi_{0,m} = \mathbb{P}(H_1 = 0)$ and $\pi_{1,m} = \mathbb{P}(H_1 = 1) = 1 - \pi_{0,m}$.

(ii) the distribution of X_1 conditionally on $H_1 = 0$ has a density $d(\cdot)$ w.r.t. the Lebesgue measure on $\mathbb{R}$ of the form $d(x) = e^{-\phi(|x|)}$ for a known function ϕ satisfying

$$\phi : \mathbb{R}^+ \to \mathbb{R} \text{ is } C^1 \text{ increasing and convex on } \mathbb{R}^+ \text{ with } \int_{\mathbb{R}} e^{-\phi(|x|)} dx = 1. \tag{A(ϕ)}$$

(iii) the distribution of X_1 conditionally on $H_1 = 1$ has a density $d_{1,m}(\cdot)$ w.r.t. the Lebesgue measure on $\mathbb{R}$ of either of the two following types:

- location: $d_{1,m}(x) = d(x - \mu_m)$, for an (unknown) location parameter $\mu_m > 0$;
- scale: $d_{1,m}(x) = d(x/\sigma_m)/\sigma_m$, for an (unknown) scale parameter $\sigma_m > 1$.

An important point in our setting is that the parameters — (τ_m, μ_m) in the location model, or (τ_m, σ_m) in the scale model — are assumed to depend on sample size m. More precisely, the parameter τ_m, called the *sparsity* parameter, is assumed to tend to infinity as m tends to infinity, which means that the unlabeled sample only contains a small, vanishing proportion of label 1. This condition is denoted (Sp). As a counterpart, the other parameter — μ_m in the location model, or σ_m in the scale model — is assumed to tend to infinity fast enough to balance sparsity. This makes the problem "just solvable" under the sparsity constraint. More precisely, our setting corresponds to the case where the Power of Bayes procedure is away from 0 and 1, and is denoted (BP).

This setting is motivated by practical situations such as source detection in astronomy or DNA copy number studies in biology, where the resolution of a measurement device increases, while the observed phenomenon is localized and has a fixed signal strength. When increasing the resolution m, the proportion $\pi_{1,m}$ of active loci decreases while the signal to noise ratio of (some of) the active loci increases (i.e., these loci are generated from a model with an increasing parameter μ_m or σ_m).

Assumption (A(ϕ)) sets a condition on $d(x) = e^{-\phi(|x|)}$ slightly stronger than "d is symmetric log concave". Namely, it also entails that $d(\cdot)$ is decreasing on $\mathbb{R}^+$. In the location model, this is essential to get a monotonic likelihood ratio, as we will see below. Also, this assumption is convenient to get expressions for tails and quantiles related to the distribution induced by $d(\cdot)$, see Appendix A. Throughout the paper, a leading example of density satisfying (A(ϕ)) is the so-called ζ-Subbotin density, $\zeta \geq 1$, defined by

$$d(x) = (L_\zeta)^{-1} e^{-|x|^\zeta/\zeta}, \text{ with } L_\zeta = \int_{-\infty}^{+\infty} e^{-|x|^\zeta/\zeta} dx, \tag{2}$$

that is, $d(x) = e^{-\phi(|x|)}$ with $\phi(u) = u^\zeta/\zeta + \log(L_\zeta)$. The particular values $\zeta = 1, 2$ give rise to the Laplace and Gaussian case, respectively. The classification problem under investigation is illustrated in Figure 1 (left panel), in the Gaussian location case.

1.3 FDR thresholding

Classically, the solution that minimizes the misclassification risk (1) is the so-called Bayes rule h_m^B that chooses the label 1 as soon as $d_{1,m}(x)/d(x)$ is larger than a specific threshold. Assuming (A(ϕ)), the likelihood ratio $d_{1,m}(x)/d(x)$ is nondecreasing in x and $|x|$ for the location and the scale model, respectively. In the location case, this comes from $\phi(|x|) - \phi(|x-\mu_m|)$ being nondecreasing in $x \in \mathbb{R}$ (because ϕ is convex increasing). In the scale case, this results from $\phi(u) - \phi(u/\sigma_m)$ being increasing in $u \in \mathbb{R}^+$ (because ϕ is convex). As a consequence, we can only focus on classification rules $\hat{h}_m(x)$ of the form $\mathbf{1}\{x \geq \hat{s}_m\}$, $\hat{s}_m \in \mathbb{R}$, for the location model, and $\mathbf{1}\{|x| \geq \hat{s}_m\}$, $\hat{s}_m \in \mathbb{R}^+$, for the scale model. Therefore, thresholding procedures are classification rules of primary interest, and the main challenge consists in choosing the threshold $\hat{s}_m$ in function of $X_1, ..., X_m$.

The FDR controlling method proposed in [2] (also called "Benjamini-Hochberg" thresholding) provides such a thresholding $\hat{s}_m$ in a very simple way once we can compute the quantile function $\overline{D}^{-1}(\cdot)$, where $\overline{D}(u) = \int_u^{+\infty} e^{-\phi(|x|)}dx$ is the (known) upper-tail cumulative distribution function of X_1 conditionally on $H_1 = 0$. In the location model, FDR thresholding is defined as follows:

Algorithm 1.1. *1. choose a nominal level $\alpha_m \in (0,1)$;*

2. *consider the order statistics of the X_k's: $X_{(1)} \geq X_{(2)} \geq ... \geq X_{(m)}$;*
3. *take the integer*
$$\hat{k} = \max\{1 \leq k \leq m : X_{(k)} \geq \overline{D}^{-1}(\alpha_m k/m)\}$$
when this set is non-empty and $\hat{k} = 1$ otherwise;
4. *use $\hat{h}_m^{FDR}(x) = \mathbf{1}\{x \geq \hat{s}_m^{FDR}\}$ for $\hat{s}_m^{FDR} = \overline{D}^{-1}(\alpha_m \hat{k}/m)$.*

For the scale model, FDR thresholding has a similar form: $\hat{h}_m^{FDR}(x) = \mathbf{1}\{|x| \geq \hat{s}_m^{FDR}\}$ for $\hat{s}_m^{FDR} = \overline{D}^{-1}(\alpha_m \hat{k}/(2m))$, where $\hat{k} = \max\{1 \leq k \leq m : |X|_{(k)} \geq \overline{D}^{-1}(\alpha_m k/(2m))\}$ ($\hat{k} = 1$ if the set is empty) and $|X|_{(1)} \geq |X|_{(2)} \geq ... \geq |X|_{(m)}$. Algorithm 1.1 is illustrated in Figure 1 (right panel), in a Gaussian location setting. Note that taking $\hat{k} = 1$ whenever the set $\{1 \leq k \leq m : X_{(k)} \geq \overline{D}^{-1}(\alpha_m k/m)\}$ is empty does not correspond to the original formulation of [2], as they choose $\hat{k} = 0$ in that case. This modification is required to tackle the "hyper-sparse" setting where $\tau_m \propto m$ (as explained in Section 6.1, it does not change the corresponding multiple testing procedure). Finally, the FDR procedure depends on a tuning parameter $\alpha_m \in (0,1)$ which should be chosen carefully, as we will explain further on.

1.4 Aim and scope of the paper

In this paper, we aim at studying the performance of FDR thresholding as a classification rule in terms of the excess risk $R_m(\hat{h}_m^{FDR}) - R_m(h_m^B)$ both in location and scale models. We investigate two types of theoretical results:

(i) Non-asymptotic oracle inequalities: prove for each (or large) m, an inequality of the form
$$R_m(\hat{h}_m^{FDR}) - R_m(h_m^B) \leq b(\phi, m, \alpha_m, \tau_m), \quad (3)$$
where $b(\phi, m, \alpha_m, \tau_m)$ is an upper-bound (depending on additional constants), which we aim to be "as small as possible".

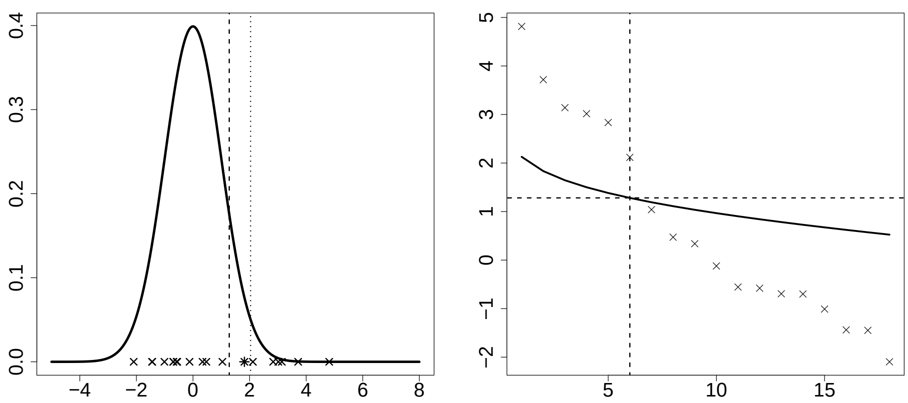

Figure 1: Left: illustration of the considered classification problem for the Gaussian location model; density of $\mathcal{N}(0,1)$ (solid line); X_k, $k=1,...,m$ (crosses); a new data point X_{m+1} to be classified (star); Bayes rule (dotted line); FDR rule $\hat{s}_m^{FDR}$ for $\alpha = 0.3$ (dashed line). Right: illustration of the FDR algorithm for $\alpha = 0.3$; $k \in \{1,...,m\} \mapsto \overline{\Phi}^{-1}(\alpha k/m)$ (solid line); $X_{(k)}$'s (crosses); $\hat{s}_m^{FDR}$ (dashed horizontal line); $\hat{k} = 6$ (dashed vertical line). Here, $\overline{\Phi}(x) = \mathbb{P}(X \geq x)$ for $X \sim \mathcal{N}(0,1)$. $m = 18$; $\mu_m = 3$; $\tau_m = 5$. For this realization, 5 labels "1" and 13 labels "0".

(ii) Convergence rate: find a sequence $(\alpha_m)_m$ such that there exists $D > 0$ such that for all $m \geq 2$,

$$R_m(\hat{h}_m^{FDR}) - R_m(h_m^B) \leq D \times R_m(h_m^B) \times \rho_m, \tag{4}$$

for a given rate $\rho_m = o(1)$.

The property (4) is called "optimal at rate ρ_m". It implies that $R_m(\hat{h}_m^{FDR}) \sim R_m(h_m^B)$, that is, $\hat{h}_m^{FDR}$ is "asymptotically optimal", as defined in [6]. However, (4) is substantially more informative because it provides a rate of convergence.

We should emphasize at this point that the trivial procedure $\hat{h}_m^0 = 0$ (which always chooses the label "0") satisfies (4) with $\rho_m = O(1)$ (under our setting (BP)). Therefore, proving (4) with $\rho_m = O(1)$ is not sufficient to get an interesting result and our goal is to obtain a rate ρ_m that tends to zero within (4). The reason for which $\hat{h}_m^0$ is already "competitive" is that we consider a sparse setting where label "0" is produced with high probability.

1.5 Overview of the paper

First, Section 2 presents a more general setting than the one of Section 1.2. Namely, the location and scale models can be seen as particular cases of a general "p-value model" after a standardization of the original X_i's into p-values p_i's. The so-obtained p-values are uniformly distributed on $(0,1)$ under the label 0 while they follow a distribution with decreasing density f_m under the label 1. Hence, procedures of primary interest (including Bayes rule) are p-value thresholding procedures, that choose the label 1 for p-values smaller than some threshold $\hat{t}_m$. Throughout the paper, we focus on this type of procedures, and any procedure $\hat{h}_m$ is identified by a threshold $\hat{t}_m$ in the notation. Translated in this "p-value world", we describe in Section 2

Bayes rule, Bayes risk, condition (BP), pFDR and FDR thresholding. pFDR thresholding, as proposed by [30], can be seen as a theoretical substitute to FDR thresholding. It is extensively used in our approach.

The fundamental results are stated in Section 3 in the general p-value model. As pFDR thresholding is much easier to study than FDR thresholding from a mathematical point of view, our approach is first to state an oracle inequality for pFDR, see Theorem 3.1, and second to use a concentration argument of the FDR threshold around the pFDR threshold to obtain an oracle inequality of the form (3), see Theorem 3.2. At this point, the bounds involve quantities which are not written under an explicit form, and which depend on the density f_m of the p-values corresponding to the label 1.

The particular case where f_m comes from a location or a scale model is investigated in Section 4. For this, an important property is that under (A(ϕ)), the upper-tail distribution function $\overline{D}(\cdot)$ and the quantile function $\overline{D}^{-1}(\cdot)$ can be bounded in function of $d(\cdot)$, ϕ, ϕ' and ϕ^{-1}, see Appendix A. By using this property, we derive from Theorems 3.1 and 3.2 several inequalities of the form (3) and (4). In particular, in the sparsity regime $\tau_m = m^{\beta}$, $0 < \beta \leq 1$, and for a ζ-Subbotin density given by (2), we derive that the FDR threshold $\hat{t}_m^{FDR}$ at level α_m is asymptotically optimal (under (BP) and (Sp)) in either of the two following cases:

- for the location model, $\zeta > 1$, if $\alpha_m \to 0$ and $\log \alpha_m = o\left((\log m)^{1-1/\zeta}\right)$;
- for the scale model, $\zeta \geq 1$, if $\alpha_m \to 0$ and $\log \alpha_m = o(\log m)$.

Furthermore, choosing $\alpha_m \propto 1/(\log m)^{1-1/\zeta}$ (location) or $\alpha_m \propto 1/(\log m)$ (scale) provides a convergence rate $\rho_m = 1/(\log m)^{1-1/\zeta}$ (location) or $\rho_m = 1/(\log m)$ (scale), respectively.

At this point, one can argue that the latter convergence results are not fully satisfactory: first, these results do not provide an explicit choice for α_m for a given finite value of m. Second, the rate of convergence ρ_m being rather slow, we can legitimately ask whether a faster rate can be obtained. Third, we should check numerically that FDR thresholding does significantly better than null thresholding for a moderately large m.

First, we address the choice of α_m by carefully studying Bayes thresholding and how it is related to pFDR thresholding, see Sections 2.5, 4.1 and 5.2. More precisely, let us consider the sparsity regime $\tau_m = m^{\beta}$, $\beta \in [\beta_-, \beta_+]$ for $0 < \beta_- < \beta_+ \leq 1$. Also, assume that the power C_m of Bayes rule lies in the range $[C_-, C_+]$ for $0 < C_- < C_+ < 1$. Then, our recommendation is to choose $\beta_0 = (\beta_- + \beta_+)/2$, $C_0 = (C_- + C_+)/2$ and

$$\alpha_m^{loc}(\beta_0, C_0) = \left\{ 1 + \frac{C_0}{d(\overline{D}^{-1}(C_0))} (\zeta \beta_0 \log m)^{1-1/\zeta} \right\}^{-1} \quad \text{for the location model, } \zeta > 1; \tag{5}$$

$$\alpha_m^{sc}(\beta_0, C_0) = \left\{ 1 + \frac{C_0/2}{\overline{D}^{-1}(C_0/2) d(\overline{D}^{-1}(C_0/2))} \zeta \beta_0 \log m \right\}^{-1} \quad \text{for the scale model, } \zeta \geq 1. \tag{6}$$

In particular, the cases $\zeta = 1, 2$ give rise to the following choices:

$$\alpha_m^{locG}(\beta_0, C_0) = \left\{1 + C_0\, e^{z_0^2/2}\sqrt{4\pi\beta_0 \log m}\right\}^{-1} \qquad \text{(Gaussian location, } \zeta = 2\text{);}$$

$$\alpha_m^{scG}(\beta_0, C_0) = \left\{1 + C_0\beta_0\sqrt{2\pi}\, e^{(z_0')^2/2}(z_0')^{-1}\log m\right\}^{-1} \qquad \text{(Gaussian scale, } \zeta = 2\text{);}$$

$$\alpha_m^{scL}(\beta_0, C_0) = \left\{1 + \beta_0(\log(1/C_0))^{-1}\log m\right\}^{-1} \qquad \text{(Laplace scale, } \zeta = 1\text{),}$$

where z_0 and z_0' denote the quantiles of order $1 - C_0$ and $1 - C_0/2$ of a standard Gaussian variable. More specifically, α_m given by either (5) or (6), denoted $\alpha_m^{\infty}(\beta_0, C_0)$ for short, is derived as an equivalent as m tends to infinity of a quantity $\alpha_m^{opt}(\beta_0, C_0)$ that enjoys some optimality property for the pFDR threshold when the model parameters are (β_0, C_0), see Sections 2.5 and 4.1. While $\alpha_m^{opt}(\beta_0, C_0)$ and $\alpha_m^{\infty}(\beta_0, C_0)$ behave similarly for large m (say, $m \geq 1000$), it is better to use $\alpha_m^{opt}(\beta_0, C_0)$ for small values of m (say $m \leq 100$). However, the level $\alpha_m^{opt}(\beta_0, C_0)$ has a less explicit expression and should be computed numerically, see Section 5.2.

Second, to address the rate issue, we provide a lower bound for the Laplace scale model in Section 4.4. More precisely, we show in that case that the rate of convergence of pFDR thresholding cannot be faster than $1/(\log m)$ for several values of β at a time (see Corollary 4.6). This means that the rate derived by our methodology is the correct one (at least for the pFDR and in the Laplace case).

Third, in Section 5, the performance of FDR thresholding (choosing α_m as suggested above) is evaluated numerically and compared to null thresholding, for several values of m and ζ. We show that the excess risk of the FDR is much smaller than the one of null thresholding for a remarkably wide range of values for β and several m. This illustrates the adaptation of FDR procedure w.r.t. the unknown sparsity regime. Also, for comparison, we show that choosing α_m fixed with m (say, $\alpha_m \equiv 0.05$) can lead to higher FDR thresholding excess risk for some values of m.

Finally, let us note that while our assumptions will exclude the case $\zeta = 1$ in the location model, our methodology can be adapted to some extent to this particular case, see Section 6.4.

1.6 Relation to previous work

First, Theorem 5.3 of [6] showed in the Gaussian scale model that FDR thresholding is asymptotically optimal, i.e., $\rho_m = o(1)$ in (4). They also found the sufficient condition $\alpha_m \to 0$ and $\log \alpha_m = o(\log m)$ (which corroborates our condition in this particular model). Our results substantially extend this work to location and scale models using symmetric log-concave densities, such as Subbotin density (2). Additionally, we also provide finite sample results with an explicit convergence rate ρ_m and a choice of α_m supported both by theory and numerical experiments. Another advantage of our approach is that our proofs appear substantially shorter and simplified.

Second, in [1] and [10], for the Gaussian location model and the Laplace scale model, respectively, it is proved that FDR thresholding is asymptotically minimax for estimating the parameter of interest (which roughly correspond to μ_m and σ_m, respectively) over specific sparsity classes (Theorem 1.1 in [1] and Theorem 1.3 in [10]). We can legitimately ask whether such a property holds in our classification framework. It would correspond to the following

property: under (BP) and (Sp), for any $\beta_0 \in (0,1)$,

$$\sup_{\beta\in[\beta_0,1]} \left\{ \frac{R_m(\hat{t}_m^{FDR})}{R_m^\star} \right\} = 1 + o(\rho_m), \tag{7}$$

with $\rho_m = o(1)$, where $R_m^\star = \inf_{\hat{t}'_m} \sup_{\beta\in[\beta_0,1]}\{R_m(\hat{t}'_m)\}$ is the minimax risk, where the infimum is taken over the set of thresholds that can be written as measurable functions of the p-values. As $R_m^\star \geq R_m(t_m^B)$ for any $\beta \in [\beta_0, 1]$, (4) implies that FDR thresholding satisfies (7), with an additional explicit rate ρ_m. However, a more interesting (but possibly more challenging) task would be to show (7) in terms of relative excess risk, as discussed in Section 6.5.

Third, the way the model parameters depend on m in our setting differs from [9] and [1]. These studies investigate detection and estimation problems, respectively, in the Gaussian location model. Their setting corresponds to the case where C_m tends to zero. This assumption is not relevant in the present classification setting, because it entails that null thresholding is asymptotically optimal (see Remark 4.2 in Section 4). In the present paper, we therefore focus on sparsity regimes where C_m remains bounded away from 0, that is, in regimes where we can hope to improve substantially over null thresholding.

Finally, let us emphasize that the classification setting described in Section 1.2 is connected to machine learning theory: namely, to Learning from Positive and Unlabeled Examples (LPUE) or Semi-Supervised Novelty Detection (SSND), see [4]. In that paper, the distribution under the label "0" is unknown but we have at hand a large sample following this distribution ("nominal" sample). The goal is to recover the labels from the nominal sample and the "contaminated" sample $X_1, ..., X_m$. However, [4] uses the Neyman-Pearson criterion, not the mis-classification risk.

2 General setting

2.1 p-value model

Let $(p_i, H_i) \in [0,1] \times \{0,1\}$, $1 \leq i \leq m$, be m i.i.d. variables. The distribution of (p_1, H_1) is assumed to belong to a specific subset of distributions on $[0,1] \times \{0,1\}$, which is defined as follows:

(i) the distribution of H_1 is such that the (unknown) mixture parameter $\tau_m = \pi_{0,m}/\pi_{1,m}$ satisfies $\tau_m > 1$, where $\pi_{0,m} = \mathbb{P}(H_1 = 0)$ and $\pi_{1,m} = \mathbb{P}(H_1 = 1) = 1 - \pi_{0,m}$;

(ii) the distribution of p_1 conditionally on $H_1 = 0$ is uniform on $(0,1)$;

(iii) the distribution of p_1 conditionally on $H_1 = 1$ has a c.d.f. F_m satisfying

$$\begin{array}{l} F_m \text{ is continuous increasing on } [0,1] \text{ and differentiable on } (0,1), \\ f_m = F'_m \text{ is continuous decreasing with } f_m(0^+) > \tau_m > f_m(1^-). \end{array} \tag{A(F_m, τ_m)}$$

This way, we obtain a family of i.i.d. p-values, where each p-value has a marginal distribution following the mixture model:

$$p_i \sim \pi_{0,m} U(0,1) + \pi_{1,m} F_m. \tag{8}$$

The model (8) is classical in the multiple testing literature and is usually called the "two-groups mixture model". It has been widely used since its introduction by Efron et al. (2001) [12], see for instance [30, 18, 11].

The models presented in Section 1.2 are particular instances of this p-value model. In the scale model, we apply the standardization $p_i = 2\overline{D}(|X_i|)$, which yields $F_m(t) = 2\overline{D}(\overline{D}^{-1}(t/2)/\sigma_m)$. We can check that if ϕ satisfies (A(ϕ)), then $F_m(t) = 2\overline{D}(\overline{D}^{-1}(t/2)/\sigma_m)$ satisfies (A(F_m, τ_m)), with $f_m(0^+) = +\infty$ and $f_m(1^-) < 1$, see Section 8.1. In the location model, we let $p_i = \overline{D}(X_i)$, which yields $F_m(t) = \overline{D}(\overline{D}^{-1}(t) - \mu_m)$. Here, (A($\phi$)) is not sufficient to ensure that $f_m = F'_m$ is decreasing, e.g. in the Laplace location model where $\phi(u) = u + \log 2$, f_m is only non-increasing. We will thus use the following additional assumption on ϕ for the location case

$$\phi \text{ satisfies (A}(\phi)\text{) and } \phi' \text{ is increasing on } \mathbb{R}^+ \text{ with } \lim_{+\infty} \phi' = +\infty, \tag{A'(ϕ)}$$

which ensures that $F_m(t) = \overline{D}(\overline{D}^{-1}(t) - \mu_m)$ satisfies (A(F_m, τ_m)), with $f_m(0^+) = +\infty$ and $f_m(1^-) = 0$, as proved in Section 8.1. Finally, an illustration of the p-value model is given in Figure 2.

Remark 2.1. *Assuming that f_m is decreasing is convenient to ensure that Bayes procedure is unique, with an explicit expression, see below. While it excludes some cases of potential interest such as the Laplace location case ($\phi(u) = u + \log 2$), it simplifies our approach. Furthermore, note that the Laplace location case is discussed separately in Section 6.4.*

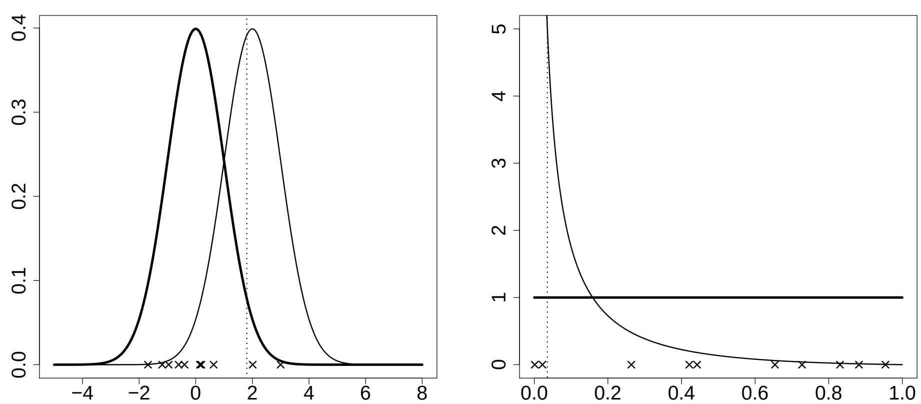

Figure 2: Left: Gaussian location model in the X_i world; density of $\mathcal{N}(0,1)$ (thick solid line); density of $\mathcal{N}(\mu_m, 1)$ (solid line); X_i, $i = 1, ..., m$ (crosses); Bayes rule (dotted line). Right: Same model and observations in the p-value world: density of $U(0,1)$ (thick solid line); density f_m (solid line); p_i, $i = 1, ..., m$ (crosses); Bayes rule (dotted line). The location parameter is $\mu_m = 2$ and the number of observations is $m = 10$.

2.2 Procedures and risk

A classification procedure is identified to a threshold $\hat{t}_m \in [0, 1]$, that is, a measurable function of the p-value family $(p_i, i \in \{1, ..., m\})$ which chooses label 1 whenever the p-value is smaller

than $\hat{t}_m$. The performance of $\hat{t}_m$ is measured via the (integrated) misclassification risk, which is defined as follows:

$$R_m(\hat{t}_m) = \mathbb{E}(\pi_{0,m}\hat{t}_m + \pi_{1,m}(1 - F_m(\hat{t}_m))). \tag{9}$$

In the particular case of a deterministic threshold $t_m \in [0,1]$, we have $R_m(t_m) = \pi_{0,m}t_m + \pi_{1,m}(1 - F_m(t_m))$.

Remark 2.2. *1. Another classical choice for the risk is the averaged mis-classification probability of $\hat{t}_m$ over the unlabeled sample itself:*

$$\begin{aligned}\widetilde{R}_m(\hat{t}_m) &= \mathbb{E}\Big(m^{-1}\sum_{i=1}^{m}\mathbf{1}\{p_i \le \hat{t}_m, H_i = 0\} + m^{-1}\sum_{i=1}^{m}\mathbf{1}\{p_i > \hat{t}_m, H_i = 1\}\Big)\\ &= m^{-1}\sum_{i=1}^{m}\mathbb{P}(p_i \le \hat{t}_m, H_i = 0) + m^{-1}\sum_{i=1}^{m}\mathbb{P}(p_i > \hat{t}_m, H_i = 1). \end{aligned} \tag{10}$$

As a matter of fact, our results also hold for this risk, as discussed in Section 6.1.

2. *Our methodology can also be easily extended to the weighted mis-classification risk, as discussed in Section 6.2. However, we have chosen to present the non-weighted case for clarity.*

2.3 Bayes procedure

An optimal thresholding is defined as any $\hat{t}_m$ satisfying

$$R_m(\hat{t}_m) = \min_{\hat{t}'_m}\{R_m(\hat{t}'_m)\}, \tag{11}$$

where the minimum is taken over all measurable functions from $[0,1]^m$ to $[0,1]$ that take as input the p-value family $(p_i, i \in \{1, ..., m\})$. By the concavity of F_m, any procedure $\hat{t}_m$ has a risk greater than its expected value, that is, $R_m(\hat{t}_m) \ge R_m(\mathbb{E}(\hat{t}_m))$. As a consequence, the minimum in (11) can be taken only over the deterministic threshold $t'_m \in [0,1]$, that is, $\min_{\hat{t}'_m}\{R_m(\hat{t}'_m)\}$ is equal to $\min_{t'_m \in [0,1]}\{R_m(t'_m)\}$. Assuming (A($F_m, \tau_m$)), the latter optimization problem has a unique solution.

Lemma 2.3. *Under Assumption (A(F_m, τ_m)), the minimum of $t \in [0,1] \mapsto R_m(t)$ exists, is unique and is given by*

$$t_m^B = f_m^{-1}(\tau_m) \in (0,1). \tag{12}$$

The threshold t_m^B is called *Bayes threshold* and $R_m(t_m^B)$ is called *Bayes risk*. Bayes threshold is unknown because it depends on τ_m and on the data distribution f_m.

2.4 Assumptions on Bayes power and Sparsity

Under Assumption (A(F_m, τ_m)), let us denote the power of Bayes procedure by

$$C_m = F_m(t_m^B) \in (0,1). \tag{13}$$

In our setting, we will typically assume that the signal is sparse while the power C_m of Bayes procedure remains away from 0 or 1:

$$\exists (C_-, C_+) \text{ s.t. } \forall m \geq 2, \;\; 0 < C_- \leq C_m \leq C_+ < 1; \tag{BP}$$

$$(\tau_m)_m \text{ is such that } \tau_m \to +\infty \text{ as } m \to +\infty. \tag{Sp}$$

First note that Assumption (Sp) is very weak: it is required as soon as we assume some sparsity in the data. As a typical instance, $\tau_m = m^\beta$ satisfies (Sp), for any $\beta > 0$. Next, Assumption (BP) means that the best procedure is able to detect a "moderate" amount of signal. In [6], a slightly stronger assumption has been introduced:

$$\exists C \in (0,1) \text{ s.t. } C_m \to C \text{ as } m \text{ tends to infinity}, \tag{VD}$$

which is referred to as "the verge of detectability" [1]. Condition (BP) encompasses (VD) while (BP) is more suitable than (VD) to obtain non-asymptotic results. Hence, (BP) will be used throughout the paper.

In the particular case of location and scale models, (BP) is equivalent to "$\overline{D}^{-1}(t_m^B) - \mu_m$ is bounded" in the location model and to "$\overline{D}^{-1}(t_m^B/2)/\sigma_m$ is bounded away from 0 and ∞" in the scale model, respectively. Moreover, while the original parameters of the model are (θ_m, τ_m) (for $\theta_m = \mu_m$ or σ_m), the model can be parametrized in function of (C_m, τ_m) by using (12) and (13). Interestingly, the c.d.f. F_m has the following interpretation w.r.t. the parameters (C_m, τ_m): among the family of curves $\{\overline{D}(\overline{D}^{-1}(\cdot) - \mu)\}_{\mu \in \mathbb{R}}$ in the location model (or $\{2\overline{D}(\overline{D}^{-1}(\cdot/2)/\sigma)\}_{\sigma > 1}$ in the scale model), $F_m(\cdot)$ is the unique curve such that the pre-image of C_m has a tangent of slope τ_m, that is, $f_m(F_m^{-1}(C_m)) = \tau_m$. This is illustrated in Figure 3 for the Laplace scale model. In this case, $\overline{D}(x) = d(x) = e^{-x}/2$ for $x \geq 0$ and thus $F_m(t) = t^{1/\sigma_m}$, so that the family of curves is simply $\{t \mapsto t^{1/\sigma}\}_{\sigma > 1}$.

2.5 pFDR thresholding

In this section, we introduce pFDR thresholding, which can be seen as a theoretical (oracle) substitute for FDR thresholding. The pFDR will be useful in our analysis because it is much easier to study than the FDR. As introduced by [30], the positive false discovery rate is defined as

$$\text{pFDR}_m(t) = \frac{\pi_{0,m} t}{G_m(t)} = \mathbb{P}(H_i = 0 \mid p_i \leq t),$$

for any $t \in (0,1)$ and $G_m(t) = \pi_{0,m} t + (1 - \pi_{0,m}) F_m(t)$. Under Assumption (A($F_m, \tau_m$)), the function $\Psi_m : t \in (0,1) \mapsto F_m(t)/t$ is decreasing from $f_m(0^+)$ to 1, with $f_m(0^+) \in (1, +\infty]$. Hence, $\text{pFDR}_m(\cdot)$ is increasing from $(1 + f_m(0^+)/\tau_m)^{-1}$ to $\pi_{0,m}$ and the following result holds.

Lemma 2.4. *Assume (A(F_m, τ_m)) and $\alpha_m \in ((1 + f_m(0^+)/\tau_m)^{-1}, \pi_{0,m})$. Then the equation $pFDR_m(t) = \alpha_m$ has a unique solution $t = t_m^\star(\alpha_m) \in (0,1)$, given by*

$$t_m^\star(\alpha_m) = \Psi_m^{-1}(q_m \tau_m), \tag{14}$$

for $q_m = \alpha_m^{-1} - 1 > 0$ and $\Psi_m(t) = F_m(t)/t$.

[1]However, we emphasize that (VD) does not refer to the so-called "detection" problem, as investigated in [9, 20] for instance.

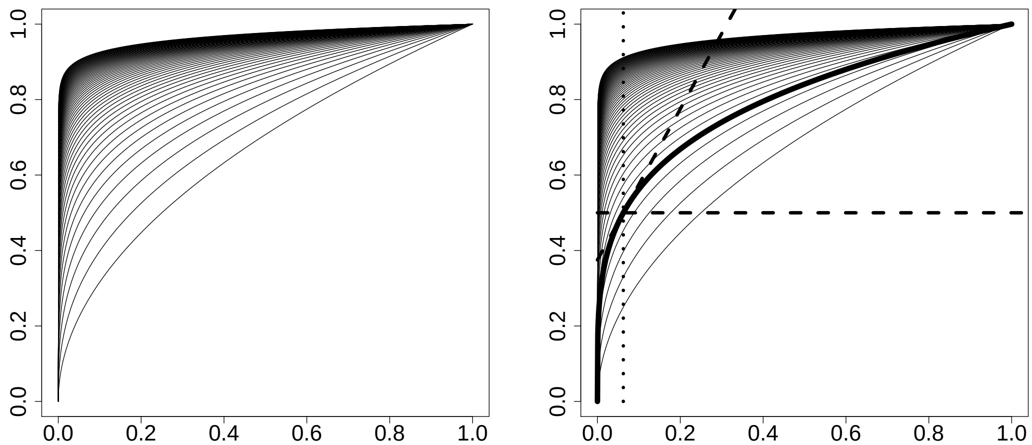

Figure 3: Left: plot of the family of curves $\{t \mapsto t^{1/(2+j/2)}\}_{j=0,\ldots,56}$ (thin solid curves). Right: choice (thick solid curve) within the family of curves $\{t \mapsto t^{1/\sigma}\}_{\sigma>1}$ that fulfills (12) and (13) for $C_m = 1/2$ (given by the dashed horizontal line) and $\tau_m = 2$ (slope of the dashed oblique line). This gives $\sigma_m \simeq 4$. Bayes threshold t_m^B is given by the dotted vertical line.

The threshold $t_m^\star(\alpha_m)$ is called the *pFDR threshold* at level α_m. The pFDR threshold is unknown because it depends on τ_m and on the distribution of the data. However, its interest lies in that it is close to the FDR threshold which is observable, as we will see in Section 3. For short, $t_m^\star(\alpha_m)$ will be denoted by $t_m^\star$ when not ambiguous.

Let us discuss the condition $\alpha_m \in ((1 + f_m(0^+)/\tau_m)^{-1}, \pi_{0,m})$ in Lemma 2.4. First, as $\pi_{0,m} > 1/2$ (because $\tau_m > 1$) and α_m will be taken smaller than $1/2$ in the sequel, we will always have $\alpha_m < \pi_{0,m}$. Second, when $f_m(0^+) = +\infty$ (for a fixed m), we have $(1 + f_m(0^+)/\tau_m)^{-1} = 0$. This case corresponds to the so-called "non-critical" case, see [8]. It is satisfied in the location and scale models considered in Section 4.

Next, Lemma 2.4 shows that the quantity $q_m = \alpha_m^{-1} - 1 > 0$ is a quantity of interest. As $\alpha_m = (1 + q_m)^{-1}$, considering α_m or q_m is equivalent.

Definition 2.5. *For each $\alpha_m \in (0, 1)$, the corresponding quantity $q_m = \alpha_m^{-1} - 1 > 0$ is called the recovery parameter (associated to α_m).*

Since we would like to have $t_m^\star = \Psi_m^{-1}(q_m\tau_m)$ close to $t_m^B = f_m^{-1}(\tau_m)$, the recovery parameter can be interpreted has a correction factor that cancels the difference between $\Psi_m(t) = F_m(t)/t$ and $f_m(t) = F_m'(t)$. In the sequel, we will always consider $q_m \geq 1$ (that is, $\alpha_m \leq 1/2$), because choosing $q_m \leq q^+ < 1$ (or equivalently $\alpha_m \geq \alpha_- > 1/2$) is always sub-optimal, see Appendix B. Clearly, the best choice for the recovery parameter is such that $t_m^\star = t_m^B$, that is,

$$q_m^{opt} = \tau_m^{-1}\Psi_m(f_m^{-1}(\tau_m)) = \frac{C_m}{\tau_m t_m^B}, \tag{15}$$

which is an unknown quantity, called the *optimal recovery parameter*. Note that from the concavity of F_m, we have $\Psi_m(t) \geq f_m(t)$ and thus $q_m^{opt} \geq 1$. As an illustration, for the Laplace

scale model, we have $\sigma_m f_m(t) = \Psi_m(t)$ and thus the optimal recovery parameter is $q_m^{opt} = \sigma_m$. This is represented in the left panel of Figure 4; the pFDR threshold for $q_m = 1$ is the point t where the line between $(0, 0)$ and $(t, F_m(t))$ is parallel to the tangent of F_m at t_m^B. In the right panel of Figure 4, the same representation is given for $G_m(t) = \pi_{0,m}t + (1-\pi_{0,m})F_m(t)$. Hence, the slopes are transformed via $u \mapsto \pi_{0,m} + (1-\pi_{0,m})u$. Note that, by definition, the pFDR threshold at level α_m is such that $G_m(t)/t = \pi_{0,m}/\alpha_m$, or, equivalently, $G_m(t)/t = \pi_{0,m}(q_m + 1)$.

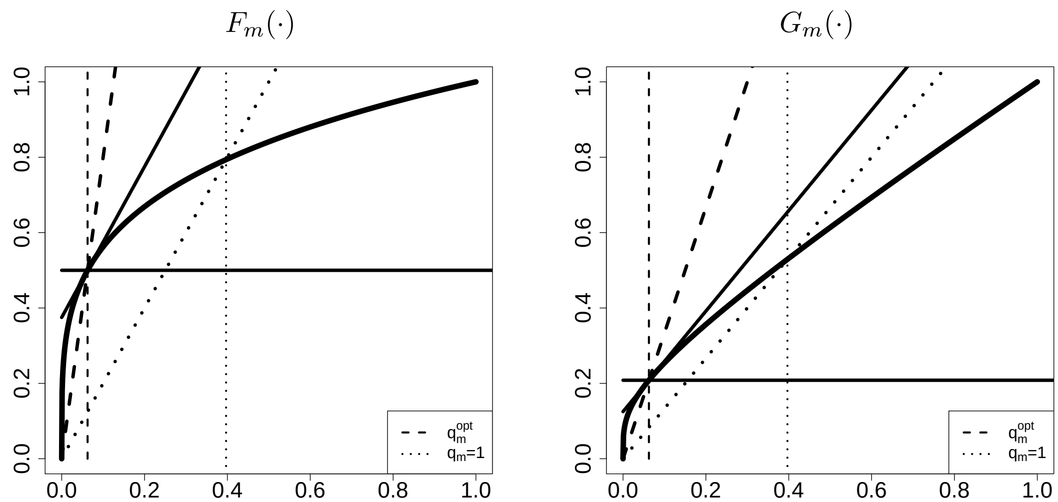

Figure 4: Recovering Bayes risk with pFDR in Laplace scale model. Left: plot of F_m (thick solid line); $C_m = 1/2$ (Y-coordinate of the horizontal solid line) and $\tau_m = 2$ (slope of the oblique solid straight line). Right: plot of $G_m(t) = \pi_{0,m}t + (1-\pi_{0,m})F_m(t)$ (thick solid line); $G_m(t_m^B) = 2t_m^B/3 + C_m/3$ is given by the Y-coordinate of the horizontal solid line and $2\pi_{0,m} = 4/3$ is the slope of the oblique solid straight line. On both pictures, pFDR thresholding is represented for $q_m = 1$ (i.e. $\alpha_m = 1/2$) (dotted) and for the optimal recovery parameter $q_m = \sigma_m \simeq 4$ (i.e. $\alpha_m \simeq 1/5$) (dashed).

2.6 FDR thresholding

The *FDR threshold* has been introduced in [2] by Benjamini and Hochberg. As noted later on by may authors (see, e.g., [17, 21]), it can be expressed as a function of the empirical c.d.f. $\widehat{\mathbb{G}}_m$ of the p-values in the following way. For any $\alpha_m \in (0, 1)$ let us define

$$\hat{t}_m^{BH}(\alpha_m) = \max\{t \in [0, 1] \, : \, \widehat{\mathbb{G}}_m(t) \geq t/\alpha_m\}. \tag{16}$$

We simply denote $\hat{t}_m^{BH}(\alpha_m)$ by $\hat{t}_m^{BH}$ when not ambiguous. Classically, this implies that $t = \hat{t}_m^{BH}$ solves the equation $\widehat{\mathbb{G}}_m(t) = t/\alpha_m$ (this can be easily shown by using (16) together with the fact that $\widehat{\mathbb{G}}_m(\cdot)$ is a non-decreasing function). Hence, according to Lemma 2.4, $\hat{t}_m^{BH}$ can be seen as an empirical substitute of the pFDR threshold at level $\alpha_m\pi_{0,m}$, in which the theoretical c.d.f. $G_m(t) = \pi_{0,m}t + \pi_{1,m}F_m(t)$ of the p-values has been replaced by the empirical c.d.f. $\widehat{\mathbb{G}}_m$ of the p-values.

Next, we would like to make the following important points about $\hat{t}_m^{BH}$: first, (16) only involves observable quantities (once α_m has been chosen), so that the threshold $\hat{t}_m^{BH}$ only depends on the data. This is further illustrated on the left panel of Figure 5. Second, let us recall the original (equivalent) definition of [2], which makes $\hat{t}_m^{BH}$ very simple to use in practice: considering the order statistics $0 = p_{(0)} \leq p_{(1)} \leq ... \leq p_{(m)}$ of the p-value family, we can write $\hat{t}_m^{BH} = \alpha_m \hat{k}_m^{BH}/m$, where $\hat{k}_m^{BH} = \max\{k \in \{0,1,...,m\} \ : \ p_{(k)} \leq \alpha_m k/m\}$. Third, for technical reasons, we chose to modify the value of $\hat{t}_m^{BH}$ in the special case where $\hat{t}_m^{BH} = 0$. When $\hat{t}_m^{BH} = 0$, we simply replace the threshold by the so-called Bonferroni threshold α_m/m.

Definition 2.6. *The FDR threshold at level α_m is defined by*

$$\hat{t}_m^{FDR} = \hat{t}_m^{BH} \vee (\alpha_m/m), \tag{17}$$

where $\hat{t}_m^{BH}$ is defined by (16).

This modification allows to deal with the "hypersparse" case $\tau_m \propto m$, as we will see later on. The threshold $\hat{t}_m^{FDR}$ is the one that we use throughout this paper. However, note that we do not need to perform this modification when considering the risk $\widetilde{R}_m$ defined in (10) instead of R_m, see discussion in Section 6.1.

Finally, we easily check that (17) and Algorithm 1.1 lead to the same classification procedure in the special case where the p-values come from a location model (obviously, the same holds for a scale model).

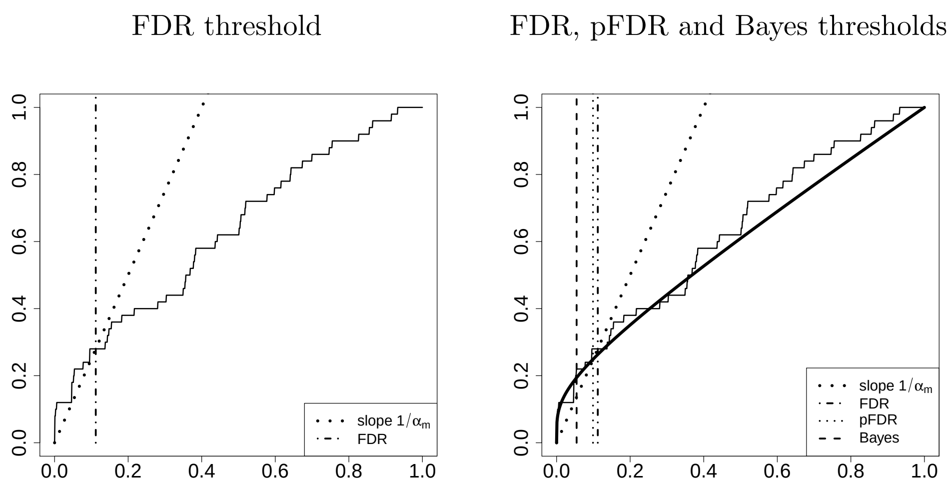

Figure 5: Left: illustration of the FDR threshold (17): e.c.d.f. of the p-value (solid line), line of slope $1/\alpha_m$ (dotted line), FDR threshold at level α_m (X-coordinate of the vertical dashed dotted line). Right: illustration of the FDR threshold as an empirical surrogate for the pFDR threshold; compared to the left picture, we added the pFDR threshold at level $\alpha_m \pi_{0,m}$ (dotted vertical line) and Bayes threshold (dashed vertical line). In both panels, we consider the Laplace scale model with $C_m = 0.5$; $m = 50$; $\beta = 0.2$; $\tau_m = m^\beta$; $\alpha_m = 0.4$.

3 General results

Choosing q_m instead of q_m^{opt} in pFDR thresholding induces some excess risk. Our first main result aims at quantifying the latter. Remember that a threshold $\hat{t}_m$ is said to be asymptotically optimal if $R_m(\hat{t}_m) \sim R_m(t_m^B)$ as m tends to infinity.

Theorem 3.1. *Assume (A(F_m, τ_m)) and consider the pFDR threshold $t_m^\star$ at a level $\alpha_m \in ((1 + f_m(0^+)/\tau_m)^{-1}, \pi_{0,m})$ corresponding to a recovery parameter $q_m = \alpha_m^{-1} - 1$. Consider $q_m^{opt} \geq 1$ the optimal recovery parameter given by (15). Then the following holds:*

(i) if $\alpha_m \leq 1/2$, we have for any $m \geq 2$,

$$R_m(t_m^\star) - R_m(t_m^B) \leq \pi_{1,m}\{(C_m/q_m - C_m/q_m^{opt}) \vee \gamma_m\}, \tag{18}$$

where we let $\gamma_m = (C_m - F_m(\Psi_m^{-1}(q_m\tau_m)))_+$. In particular, under (BP), if $\alpha_m \to 0$ and $\gamma_m \to 0$, the pFDR threshold $t_m^\star$ is asymptotically optimal at rate $\alpha_m + \gamma_m$.

(ii) we have for any $m \geq 2$,

$$\frac{R_m(t_m^\star)}{R_m(t_m^B)} \geq \frac{\pi_{1,m}}{R_m(t_m^B)} \left(1 - (1 - q_m^{-1})_+ F_m(q_m^{-1}\tau_m^{-1})\right). \tag{19}$$

In particular, under (BP), if $R_m(t_m^B) \sim \pi_{1,m}(1 - C_m)$ and if

$$\liminf_m \left\{ \frac{1 - (1 - q_m^{-1})_+ F_m(q_m^{-1}\tau_m^{-1})}{1 - C_m} \right\} > 1, \tag{20}$$

$t_m^\star$ is not asymptotically optimal.

Theorem 3.1 is proved in Section 7. Assumption $\alpha_m \leq 1/2$ in Theorem 3.1 (i) allows to get C_m/q_m instead of $1/q_m$ in the RHS of (18). This assumption is not restrictive because choosing $\alpha_m > \alpha_- \geq 1/2$ never leads to an asymptotically optimal procedure, as proved in Appendix B. Also note that the RHS of (18) is equal to zero when $q_m = q_m^{opt}$, which shows that this bound is sharp in this case.

The bound (18) induces the following trade-off for choosing α_m: on the one hand, α_m has to be chosen small enough to make C_m/q_m small; on the other hand, γ_m increases as α_m decreases to zero. The lower bound (19) is useful to identify regimes of α_m that do not lead to an asymptotically optimal pFDR thresholding.

Next, we provide our second main result, which deals with FDR thresholding.

Theorem 3.2. *Let $\varepsilon \in (0, 1)$, assume (A(F_m, τ_m)) and consider the FDR threshold $\hat{t}_m^{FDR}$ at level $\alpha_m > (1 - \varepsilon)^{-1}(\pi_{0,m} + \pi_{1,m} f_m(0^+))^{-1}$. Then the following holds: for any $m \geq 2$,*

$$\begin{aligned} R_m(\hat{t}_m^{FDR}) - R_m(t_m^B) \leq{}& \pi_{1,m}\frac{\alpha_m}{1 - \alpha_m} + m^{-1}\frac{\alpha_m}{(1 - \alpha_m)^2} \\ &+ \pi_{1,m}\left\{\gamma_m' \wedge \left(\gamma_m^\varepsilon + e^{-m\varepsilon^2(\tau_m+1)^{-1}(C_m - \gamma_m^\varepsilon)/4}\right)\right\}, \end{aligned} \tag{21}$$

for $\gamma_m^\varepsilon = (C_m - F_m(\Psi_m^{-1}(q_m^\varepsilon\tau_m)))_+$ with $q_m^\varepsilon = (\alpha_m\pi_{0,m}(1 - \varepsilon))^{-1} - 1$ and $\gamma_m' = (C_m - F_m(\alpha_m/m))_+$. In particular, under (BP) and assuming $\alpha_m \to 0$,

(i) if $\tau_m/m = O(1)$, $\gamma_m^\varepsilon \to 0$ and $\forall\kappa > 0$, $e^{-\kappa m/\tau_m} = o(\gamma_m^\varepsilon)$, the FDR threshold $\hat{t}_m^{FDR}$ is asymptotically optimal at rate $\alpha_m + \gamma_m^\varepsilon$.

(ii) if $m/\tau_m \to \ell \in (0, +\infty)$ with $\gamma'_m \to 0$, the FDR threshold $\hat{t}_m^{FDR}$ is asymptotically optimal at rate $\alpha_m + \gamma'_m$.

Theorem 3.2 is proved in Section 7. The proof mainly follows the methodology of [6], but is more general and concise. The main argument for the proof is that the FDR threshold $\hat{t}_m^{FDR}(\alpha_m)$ is either well concentrated around the pFDR threshold $t_m^\star(\alpha_m \pi_{0,m})$ (as illustrated in the right panel of Figure 5) or close to the Bonferroni threshold α_m/m.

Let us comment briefly on Theorem 3.2: first, as in the pFDR case, choosing α_m such that the bound in (29) is minimal involves a tradeoff because γ_m^ε and γ'_m are quantities that increase when α_m decreases to zero. Second, let us note that items (i) and (ii) in Theorem 3.2 are intended to cover regimes where $\tau_m = m^\beta$ with $\beta \in (0,1)$ (in which FDR is close to pFDR) and the "hyper-sparse" regime where $\tau_m = m$ (in which the FDR threshold is close to the Bonferroni threshold), respectively. Finally, the bounds and convergence rates derived in Theorems 3.1 and 3.2 strongly depend on the nature of F_m. We provide in the next section a more explicit expression of the latter in the particular cases of location and scale models.

Remark 3.3 (Conservative upper-bound for γ_m)**.** *By concavity of F_m, we have $q_m\tau_m = \Psi_m(t_m^\star) \geq f_m(t_m^\star)$, which provides*

$$\gamma_m \leq C_m - F_m(f_m^{-1}(q_m\tau_m)) \in [0,1). \tag{22}$$

When f_m^{-1} is easier to use than Ψ_m^{-1}, it is tempting to use relation (22) to upper bound the excess risk in Theorems 3.1 and 3.2. However, this can inflate too much the resulting upper-bound, as we will discuss in Section 6.3 for the case of a Gaussian density (for which this results in an additional $\log\log\tau_m$ factor in the bound).

4 Application to location and scale models

4.1 Bayes risk and optimal recovery parameter

A preliminary task is to study the behavior of t_m^B, $R_m(t_m^B)$ and $q_m^{opt} = C_m/(\tau_m t_m^B)$ both in location and scale models. Although finite sample inequalities are given in Section 8.2, we only report here some resulting asymptotic relations for short. Let us define the following rates, which will be useful throughout the paper:

$$r_m^{loc} = \phi' \circ \phi^{-1}\big(\log\tau_m + \phi(|\overline{D}^{-1}(C_m)|)\big) \tag{23}$$

$$r_m^{sc} = (\mathrm{Id} \times \phi') \circ \phi^{-1}\big(\log\tau_m + \phi(\overline{D}^{-1}(C_m/2))\big), \tag{24}$$

where Id denotes the identity function, hence, $(\mathrm{Id} \times \phi')(x) = x\phi'(x)$. Under (Sp), we easily check that the rates r_m^{loc} (resp., r_m^{sc}) tend to infinity, given that ϕ satisfies (A'(ϕ)) (resp., (A(ϕ))). Table 1 provides some useful calculations for ϕ in the case where it comes from a ζ-Subbotin density. In that case, we easily derive $r_m^{loc} = (\zeta\log\tau_m + |\overline{D}^{-1}(C_m)|^\zeta)^{1-1/\zeta}$ and $r_m^{sc} = \zeta\log\tau_m + (\overline{D}^{-1}(C_m/2))^\zeta$.

Proposition 4.1. *Consider $d(x) = e^{-\phi(|x|)}$ for a function ϕ satisfying (A(ϕ)) in the scale model or (A'(ϕ)) in the location model. Let $(\tau_m, C_m) \in (1,\infty) \times (0,1)$ be the parameters of*

$d(x)$ L_ζ $d(0)$	$(L_\zeta)^{-1}e^{-\lvert x\rvert^\zeta/\zeta}$ $\int_{-\infty}^{+\infty} e^{-\lvert x\rvert^\zeta/\zeta}dx$ $(L_\zeta)^{-1}$
$\phi(u)$ $\phi'(u)$ $\phi'\circ\phi^{-1}(v)$ $\phi^{-1}(v)\times\phi'\circ\phi^{-1}(v)$ $\phi''(u)$ $\phi''(u)/(\phi'(u))^2$	$u^\zeta/\zeta+\log L_\zeta$ $u^{\zeta-1}$ $(\zeta v-\zeta\log L_\zeta)^{1-1/\zeta}$ $\zeta v-\zeta\log L_\zeta$ $(\zeta-1)u^{\zeta-2}$ $(\zeta-1)u^{-\zeta}$

Table 1: Notation and some useful calculations for the ζ-Subbotin density.

the model. Let $r_m > 0$ be equal to r_m^{loc} defined by (23) in the location model or to r_m^{sc} defined by (24) in the scale model. Then, under (BP) and (Sp), we have

$$t_m^B = O\left(R_m(t_m^B)/r_m\right) \tag{25}$$

$$R_m(t_m^B) \sim \pi_{1,m}(1-C_m). \tag{26}$$

Furthermore, for a ζ-Subbotin density (2),

$$q_m^{opt} \sim \begin{cases} \frac{C_m}{d(\overline{D}^{-1}(C_m))}(\zeta\log\tau_m)^{1-1/\zeta} & \text{for the location model, } \zeta>1 \\ \frac{C_m/2}{\overline{D}^{-1}(C_m/2)d(\overline{D}^{-1}(C_m/2))}\zeta\log\tau_m & \text{for the scale model, } \zeta\geq 1 \end{cases}. \tag{27}$$

From (25) and (26), the probability of a type I error $\pi_{0,m}t_m^B$ is always of smaller order than the probability of a type II error $\pi_{1,m}(1-C_m)$, under (BP) and (Sp). The latter has already been observed in [6] in the particular case of a Gaussian scale model. Next, for a ζ-Subbotin density and $\tau_m = m^\beta$, $0<\beta\leq 1$, (27) gives rise to the choices $\alpha_m^{loc}(\beta_0, C_0)$ and $\alpha_m^{sc}(\beta_0, C_0)$, defined by (5) and (6), respectively, which are described in the introduction of the paper.

Remark 4.2. *From (26) and since the risk of null thresholding is $R_m(0)=\pi_{1,m}$, a substantial improvement over the null threshold can only be expected in the regime where $C_m \geq C_-$, where C_- is "far" from 0.*

4.2 Finite sample oracle inequalities

The following result can be derived from Theorem 3.1 (i) and Theorem 3.2. Itis proved in Section 8.3.

Corollary 4.3. *Consider $d(x) = e^{-\phi(|x|)}$ for a function ϕ satisfying (A(ϕ)) in the scale model or (A'(ϕ)) in the location model. Let $(\tau_m, C_m) \in (1,\infty)\times(0,1)$ be the parameters of the model. Let $r_m > 0$ and $K_m > 0$ be defined as follows:*

- *in the location model, $r_m = r_m^{loc}$ defined by (23) and $K_m = d(0)$;*
- *in the scale model, $r_m = r_m^{sc}$ defined by (24) and $K_m = 2\overline{D}^{-1}(C_m/2)d(\overline{D}^{-1}(C_m/2))$.*

Let $\alpha_m \in (0, 1/2)$ and denote the corresponding recovery parameter by $q_m = \alpha_m^{-1} - 1$. Consider $q_m^{opt} \geq 1$ the optimal recovery parameter given by (15). Let $\nu \in (0,1)$. Then:

(i) The pFDR threshold $t_m^\star$ at level α_m defined by (14) satisfies that for any $m \geq 2$ such that $r_m \geq \frac{K_m}{C_m(1-\nu)}(\log(q_m/q_m^{opt}) - \log \nu)$,

$$R_m(t_m^\star) - R_m(t_m^B) \leq \pi_{1,m} \left\{ \left(\frac{C_m}{q_m} - \frac{C_m}{q_m^{opt}} \right) \vee \left(K_m \frac{\log(q_m/q_m^{opt}) - \log \nu}{r_m} \right) \right\}; \tag{28}$$

(ii) Let $\varepsilon \in (0,1)$, $D_{1,m} = -\log(\nu \pi_{0,m}(1-\varepsilon))$ and $D_{2,m} = \log(\nu^{-1} C_m \tau_m^{-1} m)$. Then the FDR threshold $\hat{t}_m^{FDR}$ at level α_m defined by (17) satisfies that, for any $a \in \{1,2\}$, for any $m \geq 2$ such that $r_m \geq \frac{K_m}{C_m(1-\nu)}(\log(\alpha_m^{-1}/q_m^{opt}) + D_{a,m})$,

$$R_m(\hat{t}_m^{FDR}) - R_m(t_m^B) \leq \pi_{1,m} \left(\frac{\alpha_m}{1-\alpha_m} + K_m \frac{(\log(\alpha_m^{-1}/q_m^{opt}) + D_{a,m})_+}{r_m} \right) + \frac{\alpha_m/m}{(1-\alpha_m)^2} + \pi_{1,m} \mathbf{1}\{a=1\} e^{-m(\tau_m+1)^{-1}\nu\varepsilon^2 C_m/4}. \tag{29}$$

Corollary 4.3 (ii) contains two distinct cases. The case $a = 1$ should be used when m/τ_m is large, because the remaining term containing the exponential becomes small (whereas $D_{1,m}$ is approximately constant). The case $a = 2$ is intended to deal with the regime where m/τ_m is not large, because $D_{2,m}$ is of the order of a constant in that case. In any case, K_m is approximately constant with m under (BP). For instance, we can choose $\varepsilon = \nu = 1/2$ to use (28) and (29).

The form of our finite sample oracle inequalities (28) and (29) is useful to derive explicit rates of convergence, as we will see in the next section. Moreover, let us mention that (28) and (29) can be used to investigate the issue of choosing α_m for pFDR/FDR thresholding, simply by minimizing these upper-bounds in α_m, after having removing the negligible remaining terms. However, since the resulting minimum is likely to depend on some artifacts coming from the proofs (constants for instance), we prefer to use the choice induced by q_m^{opt} described in Section 4.1. Let us finally mention that an exact computation of the excess risk of pFDR thresholding can be derived in the Laplace case, see Section 4.4.

4.3 Optimality with rates

Let us recall that a threshold $\hat{t}_m$ is said to be *optimal at rate* $\rho_m = o(1)$ if there exists some constant $D > 0$ such that for all $m \geq 2$,

$$R_m(\hat{t}_m) - R_m(t_m^B) \leq D\, \rho_m\, R_m(t_m^B), \tag{30}$$

and is said asymptotically optimal if $R_m(\hat{t}_m) \sim R_m(t_m^B)$. Under (BP) and (Sp), Corollary 4.3 shows that such a result holds for pFDR/FDR thresholding, with an explicit ρ_m. Furthermore, using Theorem 3.1 (ii), we can establish a necessary and sufficient condition on α_m for which pFDR thresholding is asymptotically optimal. For this, we should introduce the following additional assumption on ϕ:

$$\phi \text{ satisfies } (\mathrm{A}(\phi)),\ \phi \text{ is } C^2 \text{ on } \mathbb{R}^+ \text{ with } \phi''/(\phi')^2 \text{ non-increasing on } (0,\infty). \tag{$B(\phi)$}$$

We will also consider the following assumption, either for $\psi = \phi' \circ \phi^{-1}$ (location) or $\psi = (\mathrm{Id} \times \phi') \circ \phi^{-1}$ (scale):

$$\psi(x + o(x)) \sim \psi(x) \text{ and } \psi(x) = O(x) \text{ as } x \to +\infty. \tag{$C(\psi)$}$$

Note that assuming (BP) and (Sp), we have $r_m \sim \psi(\log \tau_m)$ under ($C(\psi)$), either for $r_m = r_m^{loc}$ and $\psi = \phi' \circ \phi^{-1}$ or for $r_m = r_m^{sc}$ and $\psi = (\mathrm{Id} \times \phi') \circ \phi^{-1}$. Also, from Table 1, when considering a ζ-Subbotin density, Assumptions ($B(\phi)$) and ($C(\psi)$) with $\psi = \phi' \circ \phi^{-1}$ and $\psi = (\mathrm{Id} \times \phi') \circ \phi^{-1}$ are all fulfilled.

Corollary 4.4. *Consider $d(x) = e^{-\phi(|x|)}$ for a function ϕ satisfying (A(ϕ)) in the scale model or (A'(ϕ)) in the location model. Let $(\tau_m, C_m) \in (1, \infty) \times (0, 1)$ be the parameters of the model. Let $r_m > 0$ and $\psi(\cdot)$ be defined as follows:*

- *in the location model, $r_m = r_m^{loc}$ defined by (23) and $\psi = \phi' \circ \phi^{-1}$;*
- *in the scale model, $r_m = r_m^{sc}$ defined by (24) and $\psi = (Id \times \phi') \circ \phi^{-1}$.*

Assume that (BP) and (Sp) hold. Consider the pFDR threshold $t_m^\star$ at a level $\alpha_m \in (0, 1)$. Consider $q_m^{opt} \geq 1$, the optimal recovery parameter given by (15). Then the following holds:

(i) The pFDR threshold $t_m^\star$ is asymptotically optimal if

$$\alpha_m \to 0 \text{ and } \log \alpha_m = o(r_m), \tag{31}$$

in which case it is optimal at rate $\rho_m = \alpha_m + (\log(\alpha_m^{-1}/q_m^{opt}))/r_m$. Additionally, if ϕ satisfies ($B(\phi)$) and ψ satisfies ($C(\psi)$), the pFDR threshold $t_m^\star$ is asymptotically optimal if and only if (31) holds.

(ii) Further assume that there exists $\lambda > 0$ such that $\psi(x) = O(e^{\lambda x})$ for $x \to +\infty$ and that the sparsity regime τ_m satisfies

$$m/\tau_m \geq (\log \tau_m)^{1+\theta} \text{ for some } \theta > 0; \quad \text{or} \quad m/\tau_m \to \ell \in (0, +\infty). \tag{32}$$

Then, the FDR threshold $\hat{t}_m^{FDR}$ at a level α_m satisfying (31) is optimal at rate $\rho_m = \alpha_m + (\log(\alpha_m^{-1}/q_m^{opt}))/r_m$.

Let us first note that the two regimes described in (32) are the same as those proposed in [6]. They cover all possible sparse scenarios when $\tau_m = m^\beta$ with $\beta \in (0, 1]$. Next, to illustrate Corollary 4.4, let us consider the case of a ζ-Subbotin density under the sparsity regime $\tau_m = m^\beta$, for a fixed β in $(0, 1]$. In this case, the optimality condition (31) has a more explicit expression, see Table 2. Corollary 4.4 implies that this condition is necessary and sufficient for pFDR optimality, and sufficient for FDR optimality. Furthermore, it implies that convergence rate of the relative excess risk is $\rho_m = \alpha_m + \frac{\log(\alpha_m^{-1}/q_m^{opt})}{(\log m)^\gamma}$ with $\gamma = 1 - \zeta^{-1}$ (resp., $\gamma = 1$) for the location (resp., scale) case. According to the order of magnitude of q_m^{opt} (see Table 2), this proves that choosing $q_m \propto (\log m)^\gamma$ yields an optimal pFDR/FDR thresholding at rate $\rho_m = 1/(\log m)^\gamma$. For instance, the latter holds for $\alpha_m^{loc}(\beta_0, C_0)$ and $\alpha_m^{sc}(\beta_0, C_0)$ defined by (5) in the location case and by (6) in the scale case, respectively.

We can legitimately ask whether the rate $\rho_m = 1/(\log m)^\gamma$ can be improved. We show in the next section that this rate is the smallest that we can obtain over a non-trivial sparsity class, for pFDR thresholding in the Laplace scale model.

Model	ζ-Subbotin location, $\zeta > 1$	ζ-Subbotin scale, $\zeta \geq 1$
$F_m(t)$	$\overline{D}(\overline{D}^{-1}(t) - \mu_m)$	$2\overline{D}(\overline{D}^{-1}(t/2)/\sigma_m)$
Sparsity τ_m	m^β	m^β
Parameter	$\mu_m \sim (\zeta\beta \log m)^{1/\zeta}$	$\sigma_m \sim (\overline{D}^{-1}(C_m/2))^{-1}(\zeta\beta\log m)^{1/\zeta}$
r_m in (23) or (24)	$r_m^{loc} \sim (\zeta\beta\log m)^{1-1/\zeta}$	$r_m^{sc} \sim \zeta\beta\log m$
Bayes threshold		
t_m^B	$\sim m^{-\beta}\frac{d(\overline{D}^{-1}(C_m))}{(\zeta\beta\log m)^{1-1/\zeta}}$	$\sim m^{-\beta}\frac{2\overline{D}^{-1}(C_m/2)d(\overline{D}^{-1}(C_m/2))}{\zeta\beta\log m}$
$R_m(t_m^B)$	$\sim m^{-\beta}(1-C_m)$	$\sim m^{-\beta}(1-C_m)$
q_m^{opt} in (15)	$\sim \frac{C_m}{d(\overline{D}^{-1}(C_m))}(\zeta\beta\log m)^{1-1/\zeta}$	$\sim \frac{C_m/2}{\overline{D}^{-1}(C_m/2)d(\overline{D}^{-1}(C_m/2))}\zeta\beta\log m$
FDR/pFDR threshold		
Optimality condition (31)	$\alpha_m \to 0, \log\alpha_m = o\left((\log m)^{1-1/\zeta}\right)$	$\alpha_m \to 0, \log\alpha_m = o\left(\log m\right)$
Rate ρ_m in (30) for $q_m \propto q_m^{opt}$	$1/(\log m)^{1-1/\zeta}$	$1/(\log m)$

Table 2: Summary of our results for a ζ-Subbotin density in the sparsity regime $\tau_m = m^\beta$, $0 < \beta \leq 1$ and under (BP).

4.4 Case of a Laplace scale model and lower bound

In the Laplace scale model, it turns out that $\Psi_m^{-1}(\cdot)$ is an explicit function ($\Psi_m(t) = F_m(t)/t = t^{\sigma_m^{-1}-1}$), so that we can investigate exact calculations for the pFDR threshold. This is useful to establish lower bounds on the excess risk of the pFDR threshold and to get a more accurate upper-bound for the FDR threshold.

Proposition 4.5. *Consider the Laplace case $\phi(x) = x + \log 2$ and the corresponding scale model with parameters $(\tau_m, C_m) \in (1,\infty)\times(0,1)$. Let $\alpha_m \in (0,1/2)$ and $q_m = \alpha_m^{-1} - 1$ be the corresponding recovery parameter.*

(i) Let $g : x \in \mathbb{R} \mapsto e^{-x} + x - 1 \in \mathbb{R}^+$. Then the pFDR threshold $t_m^\star$ at level α_m satisfies that for any $m \geq 2$,

$$R_m(t_m^\star) - R_m(t_m^B) = C_m \pi_{1,m}\left(\frac{g(\log(q_m/\sigma_m))}{\sigma_m} + \delta_m\right), \tag{33}$$

for the remaining term $\delta_m = g\left(\frac{\log(q_m/\sigma_m)}{\sigma_m - 1}\right)(q_m^{-1} - 1) + \frac{\log(q_m/\sigma_m)}{\sigma_m-1}(\sigma_m^{-1} - q_m^{-1})$.

(ii) Let $\varepsilon \in (0,1)$, $D_{1,m} = -\log(\pi_{0,m}(1-\varepsilon))$ and $D_{2,m} = \log(m/\tau_m)$. Then the FDR threshold $\hat{t}_m^{FDR}$ at level α_m satisfies that for any $a \in \{1,2\}$, for any $m \geq 2$,

$$\begin{aligned} &R_m(\hat{t}_m^{FDR}) - R_m(t_m^B) \\ &\leq \pi_{1,m}\left(\frac{\alpha_m}{1-\alpha_m} + C_m\frac{(\log(\alpha_m^{-1}/\sigma_m) + D_{a,m})_+}{\sigma_m - 1}\right) + \frac{\alpha_m/m}{(1-\alpha_m)^2} \\ &\quad + \pi_{1,m}\mathbf{1}\{a=1\}\exp\left\{-\frac{m\varepsilon^2 C_m}{4(\tau_m+1)}\frac{(1 - (\log(\alpha_m^{-1}/\sigma_m) + D_{1,m})_+)_+}{\sigma_m - 1}\right\}. \end{aligned} \tag{34}$$

Proposition 4.5 is proved in Section 8.5. Expression (33) results from direct calculations while inequality (34) relies on Theorem 3.2. As we consider the Laplace scale model, we can

easily check that the optimal recovery parameter is σ_m, that is, we have $\Psi_m^{-1}(\sigma_m \tau_m) = t_m^B$. Expression (33) gives the excess risk when choosing q_m instead of σ_m as recovery parameter in the pFDR threshold, which is proved to strongly depend on the behavior of $g(\cdot)$. Next, inequality (34) can be seen as an improvement over (29) in the special case of a Laplace scale model: while K_m/r_m^{sc} is of the same order as C_m/σ_m in that case (because $K_m = C_m \log(1/C_m)$ and $\sigma_m \sim \log \tau_m/(\log(1/C_m))$, by using Table 2 for $\zeta = 1$), the remaining terms are of smaller order in (34) and inequality (34) is true for any $m \geq 2$.

Furthermore, expression (33) entails the following lower bound.

Corollary 4.6. *Consider the Laplace scale model satisfying assumption (BP) and (Sp). Then for any $\alpha_m \in (0, 1)$ with recovery parameter $q_m = \alpha_m^{-1} - 1$, we have*

$$R_m(t_m^\star) - R_m(t_m^B) = o\left(R_m(t_m^B)/(\log \tau_m)\right) \text{ if and only if } q_m \sim \sigma_m. \tag{35}$$

In particular, for the sparsity regimes $\tau_m = m^\beta$, $\beta \in \mathcal{B}$, for any subset $\mathcal{B}$ of $(0, 1]$ containing more than two elements, we have for any sequence $(\alpha_m)_m$ with $\alpha_m \in (0, 1)$ (that does not depend of β),

$$\liminf_m \left\{ (\log m) \sup_{\beta \in \mathcal{B}} \left(\frac{R_m(t_m^\star) - R_m(t_m^B)}{R_m(t_m^B)} \right) \right\} > 0. \tag{36}$$

Corollary 4.6 is proved in Section 8.5. For the sparsity regimes $\tau_m = m^\beta$, the equivalence in (35) shows that the only way to obtain a relative excess risk of order smaller than $(\log m)^{-1}$ is to take $q_m \sim \beta \log m/(\log(1/C_m))$. This choice is not possible when β can take several values. This gives rise to the formulation (36). As a consequence, the rate obtained in Corollary 4.4 (itself coming from Corollary 4.3) may not be improved for pFDR thresholding in the particular case of a Laplace scale model.

While the calculations become significantly more difficult in the other models, we believe that the minimal rate for the relative excess risk of the pFDR is still $(\log m)^{-\gamma}$ for a ζ-Subbotin density, with $\gamma = 1 - \zeta^{-1}$ (resp. $\gamma = 1$) for the location (resp. scale) case. Also, since the FDR can be seen as a stochastic variation around the pFDR, we believe that this rate is also minimal in the case of the FDR, see also the discussion in Section 6.5.

5 Numerical experiments

In order to complement the convergence results stated above, it is of interest to study the behavior of FDR and pFDR thresholding for a small or moderate m.

5.1 Exact formula and upper-bound for the FDR risk

The pFDR threshold $t_m^\star$ can be approximated numerically, which allows us to compute $R_m(t_m^\star)$. Computing $R_m(\hat{t}_m^{FDR})$ is more complicated, because the FDR threshold $\hat{t}_m^{FDR}$ is not deterministic. However, we can avoid performing cumbersome and somewhat imprecise simulations to compute $R_m(\hat{t}_m^{FDR})$, by using the approach proposed in [15] and [25]. Using this methodology, the full distribution of $\hat{t}_m^{FDR}$ may be written as a function of the c.d.f. of the order statistics of i.i.d. uniform variables. Let for any $k \geq 0$ and for any $(t_1, ..., t_k) \in [0, 1]^k$,

$$\Psi_k(t_1, ..., t_k) = \mathbb{P}(U_{(1)} \leq t_1, ..., U_{(k)} \leq t_k),$$

where $(U_i)_{1\leq i\leq k}$ is a sequence of i.i.d. uniform variables on $(0,1)$ and with the convention $\Psi_0(\cdot)=1$. The Ψ_k's can be evaluated e.g. by using Steck's recursion (see [28], pages 366-369). Then, relation (10) in [25] entails

$$\begin{aligned} R_m(\hat{t}_m^{FDR}) = &\sum_{k=0}^{m} \binom{m}{k} R_m\left(\frac{\alpha(k\vee 1)}{m}\right) G_m(\alpha k/m)^k \\ &\times \Psi_{m-k}\big(1-G_m(\alpha m/m), ..., 1-G_m(\alpha(k+1)/m)\big), \end{aligned} \tag{37}$$

where $G_m(t)=\pi_{0,m}t+\pi_{1,m}F_m(t)$. For reasonably large m ($m\leq 10,000$ in what follows), expression (37) can be used for computing the *exact* risk of FDR thresholding $\hat{t}_m^{FDR}$ in our experiment.

For larger m, e.g., $m=10^6$, we did not undertake exact FDR risk calculations, because evaluating Ψ_k, $k\in\{1,\dots m\}$ was not feasible in practice, for two reasons. First, available algorithms are quadratic in m. Second, this calculation involved the summation of very large numbers of very small terms, making the numerical accuracy of the result questionable for very large m. Nevertheless, as $\Psi_k(t_1,...,t_k)\leq\Psi_k(t_k,...,t_k)=(t_k)^k$, we propose to replace (37) by the following upper-bound for the risk:

$$R_m(\hat{t}_m^{FDR}) \leq \sum_{k=0}^{m} \binom{m}{k} R_m\left(\frac{\alpha(k\vee 1)}{m}\right) G_m(\alpha k/m)^k (1-G_m(\alpha(k+1)/m))^{m-k}. \tag{38}$$

This upper bound can be calculated quickly and with great numerical accuracy even for large m (e.g., $m=10^6$).

5.2 Choosing α_m

By using (15) in Section 2.5, we propose to choose α_m as follows:

$$\alpha_m^{opt}(\beta_0,C_0)=\big(1+q_m^{opt}(\beta_0,C_0)\big)^{-1} \text{ with } q_m^{opt}(\beta_0,C_0)=m^{-\beta_0}C_0/F_{m,0}^{-1}(C_0), \tag{39}$$

where $F_{m,0}$ is the c.d.f. of the p-values following the alternative for the model parameters (β_0,C_0). For instance,

$$\begin{aligned} F_{m,0}^{-1}(C_0) &= \overline{\Phi}\left(\left\{\overline{\Phi}^{-1}(C_0)^2+2\beta_0\log m\right\}^{1/2}\right); && \text{(Gaussian location)} \\ F_{m,0}^{-1}(C_0) &= 2\overline{\Phi}\left(\overline{\Phi}^{-1}(C_0/2)x\right), && \text{(Gaussian scale)} \\ &\text{with } x>1 \text{ solving } 2\beta_0\log m+2\log x=(\overline{\Phi}^{-1}(C_0/2))^2(x^2-1); \\ q_m^{opt}(\beta_0,C_0) &= y>1 \text{ solves } \beta_0\log m+\log y=(y-1)\log(1/C_0), && \text{(Laplace scale)} \end{aligned}$$

where $\overline{\Phi}(z)$ denotes $\mathbb{P}(Z\geq z)$ for $Z\sim\mathcal{N}(0,1)$. From Proposition 4.1, the choice $\alpha_m^{opt}(\beta_0,C_0)$ defined by (39) is asymptotically equivalent to the explicit choice $\alpha_m^{\infty}(\beta_0,C_0)$ given by (5) and (6) in the introduction of the paper. Numerical comparisons between the pFDR and FDR risks obtained according to $\alpha_m^{opt}(\beta_0,C_0)$ and $\alpha_m^{\infty}(\beta_0,C_0)$ are provided in Section 2 of the supplementary material [22]. While $\alpha_m^{\infty}(\beta_0,C_0)$ qualitatively leads to the same results when m is large (say, $m\geq 1,000$), $\alpha_m^{opt}(\beta_0,C_0)$ is more accurate for a small m.

Finally, note that the choices $\alpha_m^{opt}(\beta_0,C_0)$ and $\alpha_m^{\infty}(\beta_0,C_0)$ are motivated by the analysis of the pFDR risk, not that of the FDR risk. Hence, it might be possible to choose a better

α_m for FDR, especially for small values of m for which pFDR and FDR are different. Because obtaining such a refinement appeared quite challenging, and as our proposed choice already performed well, we decided not to investigate this question further.

5.3 Adapting to unknown sparsity

In order to make our experiments comparable across parameter values, we quantify the quality of a thresholding procedure based on the ratio between the excess risk of this procedure to the excess risk of null thresholding:

$$\mathrm{ERR}_m(\hat{t}_m) = \frac{R_m(\hat{t}_m) - R_m(t_m^B)}{R_m(0) - R_m(t_m^B)}. \tag{40}$$

The excess risk ratio $\mathrm{ERR}_m(\hat{t}_m)$ defined by (40) is the baseline for all our experiments. The closer it is to 0, the better the corresponding classification procedure is. Figure 6 compares excess risk ratios of different procedures in the Gaussian location model: Bayes procedure with parameters ($\beta = \beta_0, C_m = C_0$) (that is, $F_{m,0}^{-1}(C_0)$), pFDR and FDR thresholding at level α for $\alpha \in \{0.05, 0.1, 0.2, \alpha_m^{opt}(\beta_0, C_0)\}$. We study the behavior of the excess risk ratio as the (unknown) true model parameters (β, C_m) vary in $[0,1]\times[0,1]$, and arbitrarily choose β_0 and C_0 as the midpoints of the corresponding intervals, i.e. $\beta_0 = 1/2$ and $C_0 = 1/2$. Colors reflect the value of the excess risk ratio. They range from white (low risk) to dark red (higher risk). Black lines represent the level set ERR $= 0.2$, that is, they delineate a region of the (β, C_m) plane in which the excess risk of the procedure under study is five times less than that of null thresholding. The number at the top left of each plot gives the fraction of configurations (β, C) for which ERR ≤ 0.2. Each column in Figure 6 corresponds to a value of $m \in \{25, 100, 10^3, 10^4, 10^6\}$. For $m = 10^6$, we did not undertake exact FDR risk calculations, but used (38) to provide an upper bound on the FDR relative risk for $m = 10^6$. We expect this bound to be conservative, and the corresponding plots are marked with $(*)$. Also note that FDR risk is expected to be well approximated by pFDR risk for such a large value of m. This is confirmed by the fact that FDR and pFDR plots at a given level α are increasingly similar as m increases.

Bayes thresholding (top line) performs well when the sparsity parameter β is correctly specified, and its performance is fairly robust to C_m. However, it performs poorly when β is misspecified, and increasingly so as m increases. The results are markedly different the other thresholding methods. pFDR and FDR thresholding are less adaptive to C_m than Bayes thresholding, but much more adaptive to the sparsity parameter β, as illustrated by the fact that the configurations with low ERR span the whole range of β, especially when $\alpha = \alpha_m^{opt}(\beta_0, C_0)$.

Another striking point is that while pFDR thresholding with fixed values of α performs fairly well for some values of m, it is outperformed by pFDR thresholding when $\alpha = \alpha_m^{opt}(\beta_0, C_0)$. This is because this choice of α is calibrated *as a function of* m. The same remark holds for FDR thresholding. Importantly, pFDR and FDR thresholding using this calibration are *increasingly* adaptive to sparsity as m increases. This corroborates the results of Section 4.3 which entail that $\mathrm{ERR}_m(t_m^\star)$ and $\mathrm{ERR}_m(\hat{t}_m^{FDR})$ are $O((\log m)^{-1/2})$.

Results for Laplace and Gaussian scale models are similar. The corresponding Figures are given in Section 2 of the supplementary material [22]. Importantly, the range of values of $\alpha_m^{opt}(\beta_0, C_0)$ differs substantially between models: from $[0.17, 0.27]$ in the Gaussian location model, to $[0.05, 0.12]$ in the Gaussian scale model and $[0.06, 0.15]$ in the Laplace scale model.

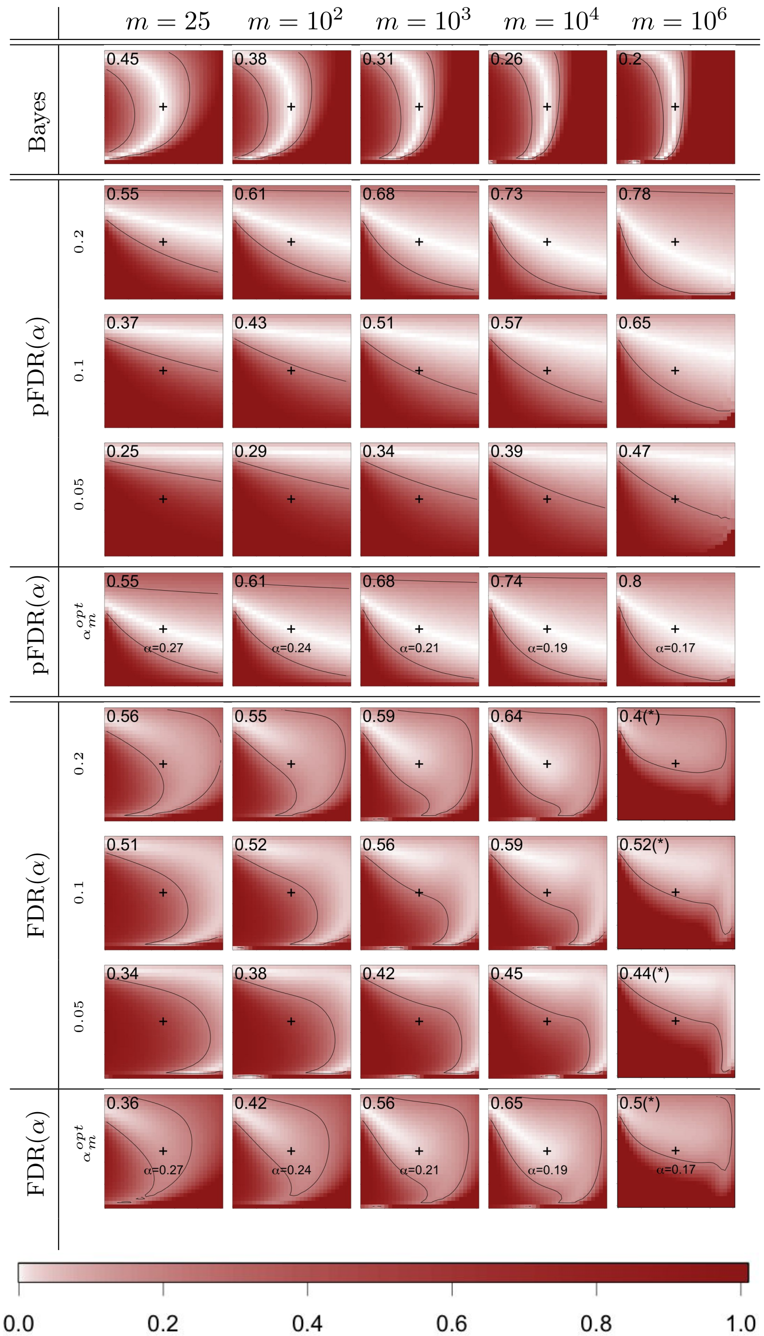

Figure 6: Adaptation to sparsity by (p)FDR thresholding in the Gaussian location model. Excess risk ratios ERR_m for various thresholding procedures (rows) and different values of m (columns). In each panel, the corresponding risk is plotted as a function of $\beta \in [0,1]$ (horizontal axis) and $C_m \in [0,1]$ (vertical axis). Colors range from white (low risk) to dark red (high risk), as indicated by the color bar at the bottom. For FDR, panels with $m = 10^6$ are marked with a star ($\star$) in order to indicate that only an upper bound on ERR was calculated. Black lines represent the level set $\mathrm{ERR} = 0.2$. The point $(\beta = \beta_0, C = C_0)$ is marked by "+". We chose $\beta_0 = 1/2$ and $C_0 = 1/2$.

5.4 Influence of the choice of parameters β_0 and C_0

In figure 6, Bayes procedure and the optimal recovery parameters are calibrated using $\beta_0 = 1/2$ and $C_0 = 1/2$. The above results show that pFDR and FDR thresholding are adaptive to the unknown sparsity, in the sense that when applied at level $\alpha_m^{opt}(\beta_0, C_0)$, they achieve a low excess risk ratio even when the true sparsity parameter β is not β_0.

In this section we discuss the influence of C_0 on the performance of pFDR and FDR thresholding at level $\alpha_m^{opt}(\beta_0, C_0)$. Figure 7 gives the ERR for Bayes, pFDR and FDR thresholding for $\beta_0 = 1/2$ and $C_0 \in \{1/4, 1/2, 3/4\}$. As expected, Bayes thresholding is quite robust to the choice of C_0, as it achieves low ERR for all values of C_0. However, as mentioned above, Bayes thresholding is quite sensitive to the specification of β, and its performance when β_0 is misspecified decreases rapidly as m increases. In contrast, pFDR thresholding at $\alpha_m^{opt}(\beta_0, C_0)$ is more sensitive to the specification of C_m, and much less to the specification of β. In particular, the region in the (β, C_m) plane for which ERR ≤ 0.2 are markedly different for $C_0 = 1/4$, $1/2$ or $3/4$, especially for small values of m. As m increases, these low-ERR regions widen and their overlap increases, making pFDR thresholding less sensitive to the specification of C_m. FDR thresholding at $\alpha_m^{opt}(\beta_0, C_0)$ achieves a reasonably low ERR over the whole range of values for β and C_m. However, the region with low ERR is smaller for smaller values of C_0. We also observe that for a given value of C_0, the region with low ERR gets bigger as m increases. We believe that this also holds for larger m, even if it cannot be deduced from the upper bound on FDR ERR that we calculated for $m = 10^6$.

Results for Laplace and Gaussian scale models are similar. The corresponding Figures are given in Section 2 of the supplementary material [22] .

6 Discussion

6.1 Extension to the risk $\widetilde{R}_m$

Our bounds are established for the misclassification risk over a new labeled data (9) and not for the misclassification risk over the unlabeled sample $\widetilde{R}_m$, defined by (10). Remember that these two risks are the same for a deterministic threshold (e.g., the pFDR threshold), but can be different for a random threshold. Hence Theorem 3.1 also holds for the risk $\widetilde{R}_m$. We can legitimately ask whether this is the case for Theorem 3.2.

As a matter of fact, we can prove that Theorem 3.2 is also true for the risk $\widetilde{R}_m$; first, for this risk, the threshold $\hat{t}_m^{FDR}$ defined by (17) has the same risk than the threshold $\hat{t}_m^{BH}$ defined by (16). This comes from the equality

$$\{1 \leq i \leq m \;:\; p_i \leq \hat{t}_m^{FDR}\} = \{1 \leq i \leq m \;:\; p_i \leq \hat{t}_m^{BH}\},$$

which can be easily checked. Hence we can work directly with $\hat{t}_m^{BH}$. Second, the bound for the type I error is the same as in (51)and can be proved similarly. Third, the proof for bounding the type II error derives essentially from the following argument, which is quite standard in the multiple testing methodology, see e.g. [13, 14, 25, 23]. Let us denote

$$\tilde{t}_m = \max\{t \in [0,1] \;:\; \alpha_m \widetilde{\mathbb{G}}_m(t) \geq t\},$$

where $\widetilde{\mathbb{G}}_m(t) = m^{-1}(1 + \sum_{i=2}^m \mathbf{1}\{p_i \leq t\})$ denotes the empirical c.d.f. of the p-values where p_1 has been replaced by 0. Then, for any realization of the p-value family, $p_1 \leq \hat{t}_m^{BH}$ is equivalent

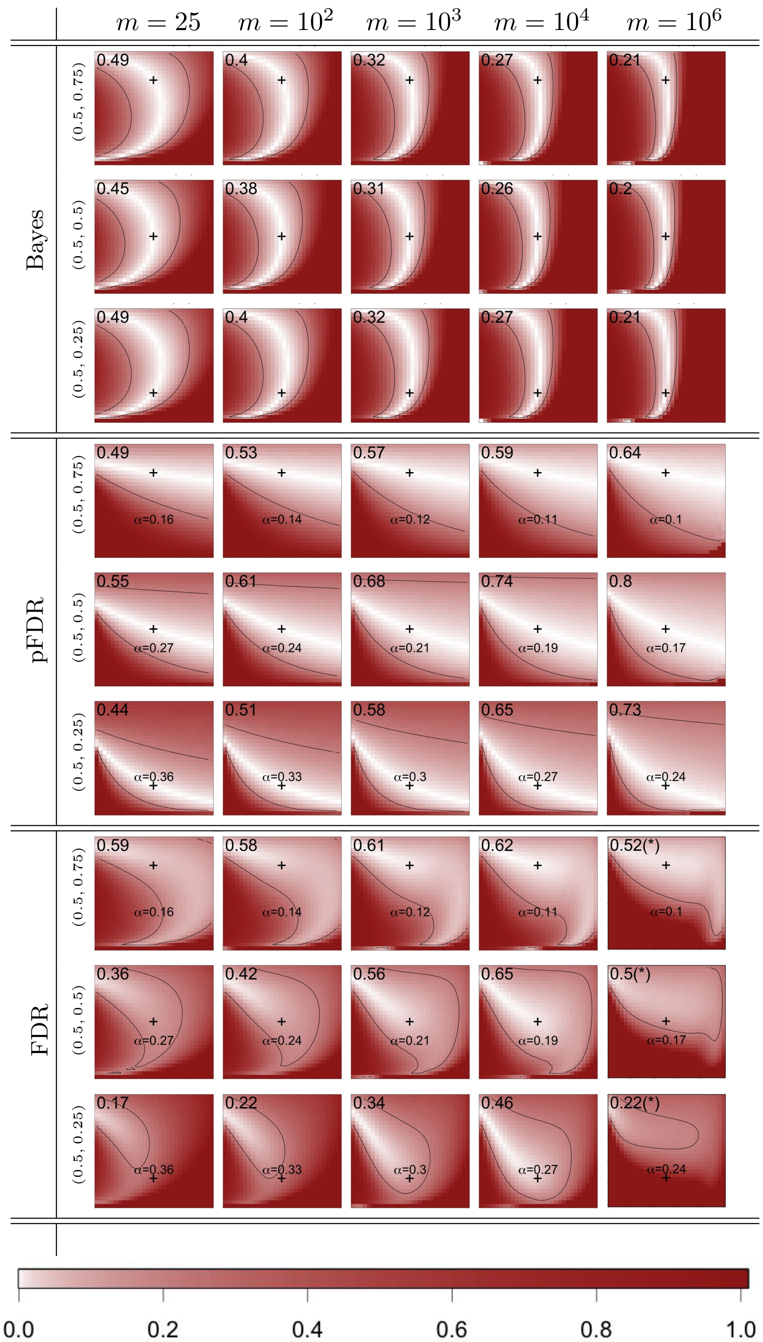

Figure 7: Excess risk ratios (ERR) of Bayes, pFDR and FDR thresholding for $m \in \{25, 100, 10^3, 10^4, 10^6\}$, $\beta_0 = 1/2$ and $C_0 \in \{0.25, 0.5, 0.75\}$. In each panel, the point $(\beta = \beta_0, C = C_0)$ is marked by "+".

to $p_1 \le \tilde{t}_m$ (see, e.g., Section 3.2 of [23]). This entails that the type II error is equal to $\pi_{1,m}(1-\mathbb{E}(F_m(\tilde{t}_m)))$ (by using the exchangeability of $(H_i, p_i)_{1\le i\le m}$). Finally, since $\tilde{t}_m \ge \hat{t}_m^{BH}$ and $\tilde{t}_m \ge \alpha_m/m$, we have $\tilde{t}_m \ge \hat{t}_m^{FDR}$. Hence $\pi_{1,m}(1-\mathbb{E}(F_m(\tilde{t}_m))) \le \pi_{1,m}(1-\mathbb{E}(F_m(\hat{t}_m^{FDR})))$ and the bounds (54) and (55) also hold for the risk $\widetilde{R}_m$.

In conclusion, all the results of Sections 3 and 4 are valid using the risk $\widetilde{R}_m$ instead of R_m.

6.2 Extension to weighted mis-classification risk

In our sparse setting, where we assume that there are many more labels "0" than labels "1", one could consider that mis-classifying a "0" is less important than mis-classifying a "1". This suggests to consider the following weighted risk:

$$R_{m,\lambda_m}(\hat{t}_m) = \mathbb{E}(\pi_{0,m}\hat{t}_m + \lambda_m \pi_{1,m}(1 - F_m(\hat{t}_m))), \tag{41}$$

for a known factor $\lambda_m \in (1, \tau_m)$. In Section 1 of the supplementary material [22], we show that all our results can be adapted to this risk. Loosely, when considering R_{m,λ_m} instead of R_m, our results hold after replacing τ_m by τ_m/λ_m and q_m by $q_m\lambda_m$, see the supplementary material [22] for precise statements.

As an illustration, let us consider here the case of a ζ-Subbotin density, $\tau_m = m^\beta$, $\beta \in (0, 1]$, $\log \lambda_m = o(\log m)$, under the (corresponding) assumptions (BP) and (Sp). We show that the optimal recovery parameter satisfies $q_m^{opt}\lambda_m \propto (\log m)^\gamma$, where $\gamma = 1-\zeta^{-1}$ and $\gamma = 1$ for the location and scale cases, respectively. Furthermore, we show that taking $q_m \propto q_m^{opt}$ leads to the optimality rate $\rho_m = (\log m)^{-\gamma}$ for the relative excess risk based on R_{m,λ_m}. While the order of q_m^{opt} is not modified when $\lambda_m \propto 1$, it may be substantially different when $\lambda_m \to \infty$. Typically, $\lambda_m \propto (\log m)^\gamma$ leads to $q_m^{opt} \propto 1$. Hence, when considering R_{m,λ_m} instead of R_m, the value of λ_m should be carefully taken into account when choosing α_m to obtain a small excess risk.

Finally, for the ζ-Subbotin density, $\tau_m = m^\beta$, $\beta \in (0, 1]$ and $\log \lambda_m = o((\log m)^\gamma)$, we show in the supplementary material [22] that a sufficient condition for FDR thresholding to be asymptotically optimal for the risk R_{m,λ_m} is to take $q_m^{-1} = O(1)$, $q_m\lambda_m \to \infty$ and $\log q_m = o\left((\log m)^\gamma\right)$. This recovers Theorem 5.3 of [6] when applied to the particular case of a Gaussian scale model (for which $\gamma = 1$).

6.3 Case of a Gaussian density

Let us consider the special case where $d(\cdot)$ is the standard Gaussian density. In that case, while Ψ_m is not easily invertible, an explicit expression can be derived for f_m^{-1}, see Table 3. By using (22) in Remark 3.3, Theorems 3.1 and 3.2 lead to explicit upper bounds for the excess risk of the pFDR/FDR. By contrast with the bounds derived in Section 4.2, they are valid for any $m \ge 2$, but the quantity "$\log(q_m/q_m^{opt})$" is replaced by "$\log q_m$" (up to constant terms). The reason is that $\gamma_m = (F_m(\Psi_m^{-1}(q_m^{opt}\tau_m)) - F_m(\Psi_m^{-1}(q_m\tau_m)))_+$ involves a variation of q_m around q_m^{opt}, while $C_m - F_m(f_m^{-1}(q_m\tau_m)) = F_m(f_m^{-1}(\tau_m)) - F_m(f_m^{-1}(q_m\tau_m))$ involves a variation of q_m around 1. When choosing $q_m \propto q_m^{opt}$, this method inflates the upper-bound by a factor $\log\log\tau_m$ w.r.t. the bounds derived in Section 4.2. Hence, we have chosen to not report these bounds in the final manuscript.

	Gaussian location	Gaussian scale
Parameter	$\mu_m = -\overline{\Phi}^{-1}(C_m) + \sqrt{(\overline{\Phi}^{-1}(C_m))^2 + 2\log\tau_m}$	$\log\tau_m + \log\sigma_m = (\overline{\Phi}^{-1}(C_m/2))^2(\sigma_m^2 - 1)/2$
$F_m(t)$	$\overline{\Phi}(\overline{\Phi}^{-1}(t) - \mu_m)$	$2\overline{\Phi}(\overline{\Phi}^{-1}(t/2)/\sigma_m)$
$f_m(t)$	$\exp(\mu_m(\overline{\Phi}^{-1}(t) - \mu_m/2))$	$\sigma_m^{-1}\exp\{(1-\sigma_m^{-2})(\overline{\Phi}^{-1}(t/2))^2/2\}$
$f_m^{-1}(u)$	$\overline{\Phi}((\log u)/\mu_m + \mu_m/2)$	$2\overline{\Phi}\big((2(\log(\sigma_m u))\sigma_m^2/(\sigma_m^2-1))^{1/2}\big)$
$F_m(f_m^{-1}(q_m\tau_m))$	$\overline{\Phi}((\log q_m)/\mu_m + \overline{\Phi}^{-1}(C_m))$	$2\overline{\Phi}\left(\left((\overline{\Phi}^{-1}(C_m/2))^2 + \frac{2\log q_m}{\sigma_m^2-1}\right)^{1/2}\right)$

Table 3: Some calculations for the Gaussian location and scale models. $\overline{\Phi}(x) = \mathbb{P}(Z \geq x)$ for $Z \sim \mathcal{N}(0,1)$; $t \in (0,1)$; $u > 0$.

6.4 Laplace location model

According to Remark 2.1, our results do not cover the case of the Laplace location model because $\phi(u) = u + \log 2$ is not strictly convex. In this case, while the optimal classification procedures are still the thresholding procedures, Bayes threshold is 0 or 1 whenever $\tau_m \leq e^{-\mu_m}$ or $\tau_m \geq e^{\mu_m}$, respectively. This can be derived from the exact expression of F_m provided in Proposition 25 of [20] (item 3). Nevertheless, Bayes threshold is still unique in $(0,1)$ as soon as the parameters (τ_m, μ_m) satisfy the constraint

$$e^{-\mu_m} < \tau_m < e^{\mu_m}. \tag{42}$$

Moreover, this entails $1/2 < C_m < 1 - e^{-\mu_m}/2$, $q_m^{opt} = C_m/(1-C_m)$ and $R_m(t_m^B) = 2\pi_{1,m}(1-C_m)$. In particular, one major difference with the cases considered in this paper is that q_m^{opt} does not tend to infinity under (BP) and (Sp). Also, we have $r_m^{loc} = 1$ as defined in (23). Under Assumption (42), Theorems 3.1 and 3.2 can be readily applied to obtain upper bounds for the excess risk of pFDR/FDR thresholding. While this proves that pFDR thresholding is still asymptotically optimal when choosing $q_m - q_m^{opt} = o(1)$, we cannot derive directly such a statement for FDR thresholding. This comes from the fact that we used a "one-sided" concentration argument while bounding the type I error. Rather, we would need a "two-sided" concentration argument, which seems feasible but maybe technical.

We have also performed numerical experiments for the Laplace location model, see Figure 3 in the supplementary material [22]. These experiments show that this model is somewhat singular: while the adaptation w.r.t. β is stronger than for the other models (ERR is even *independent of* β for pFDR thresholding), the sensitivity to the mis-specification of C_m is much higher. This behavior is in agreement with the expression of q_m^{opt} which involves C_m but not β.

6.5 Asymptotically minimax relative excess risk

Let us denote the relative excess risk $\mathcal{E}_m(\hat{t}_m) = (R_m(\hat{t}_m) - R_m(t_m^B))/R_m(t_m^B)$ and consider the sparsity range $\tau_m = m^{-\beta}$, $\beta \in \mathcal{B}$, for a subset $\mathcal{B}$ of $(0,1]$ containing at least two elements. Let us focus on the Laplace scale model. We showed in Section 4 that, under (BP), (Sp) and by taking $\alpha_m \propto (\log m)^{-1}$, there exists some constant $D > 0$ such that for $m \geq 2$,

$$\sup_{\beta\in\mathcal{B}} \left\{\mathcal{E}_m\left(\hat{t}_m^{FDR}(\alpha_m)\right)\right\} \leq \frac{D}{\log m}. \tag{43}$$

Furthermore, (36) shows that the rate in (43) is not improvable over the class of pFDR procedures using an arbitrary nominal level $\alpha_m \in (0,1)$. An interesting open problem for

future research is to determine whether there exists a procedure $\hat{t}_m$ achieving a faster rate than $(\log m)^{-1}$. We might conjecture that this is not the case, i.e., that there exists some constant $D' > 0$ such that for $m \geq 2$,

$$\inf_{\hat{t}_m} \left\{ \sup_{\beta \in \mathcal{B}} \{\mathcal{E}_m(\hat{t}_m)\} \right\} \geq \frac{D'}{\log m}, \tag{44}$$

where the infimum is taken over any thresholding procedure $\hat{t}_m : [0,1]^m \to [0,1]$ taking as input the p-value family. The latter, combined with (43), would show that FDR thresholding is asymptotically minimax in terms of relative excess risk. This would be more accurate than a result of the form (7) and is thus an interesting direction for future investigations.

6.6 Case of other FDR controlling procedures

The present paper focuses on the seminal FDR controlling procedure proposed by Benjamini and Hochberg [2], which is based on Simes' line [29]. However, many other procedures have been proved to control FDR while they proposed some refinements over [2], for instance, step-up-down procedures, see, e.g., [31, 26], procedures adaptive to $\pi_{0,m}$, see, e.g., [3, 27, 5, 14, 16], or procedures adaptive to the alternative c.d.f. F_m, see [24]. We believe that some of these procedures also have the property to be adaptive to unknown sparsity, and may outperform [2] as a classification rule. This is an interesting avenue for future research.

7 Proofs of Theorem 3.1 and Theorem 3.2

7.1 Relations for pFDR

Let us first state the following result.

Proposition 7.1. *Consider the setting and the notation of Theorem 3.1. Then we have*

1. *for any $m \geq 2$,*

$$R_m(t^\star_m) - R_m(t^B_m) = \pi_{1,m} C_m / q_m - \pi_{0,m} t^B_m + \pi_{1,m}(1 - q_m^{-1})(C_m - F_m(t^\star_m)). \tag{45}$$

2. *if $\alpha_m \leq 1/2$, we have for any $m \geq 2$,*

$$R_m(t^\star_m) - R_m(t^B_m)) \leq \pi_{1,m} C_m / q_m - \pi_{0,m} t^B_m + \pi_{1,m}(1 - q_m^{-1})\gamma_m \tag{46}$$

$$R_m(t^\star_m) - R_m(t^B_m) \leq \pi_{1,m}(C_m / q_m - \tau_m t^B_m) \vee \gamma_m. \tag{47}$$

Proof. To prove (45), we use $F_m(t^\star_m) = t^\star_m q_m \tau_m$ and $\tau_m = \pi_{0,m}/\pi_{1,m}$, to write

$$\begin{aligned} R_m(t^\star_m) - R_m(t^B_m) &= \pi_{0,m} t^\star_m - \pi_{0,m} t^B_m + \pi_{1,m}(C_m - F_m(t^\star_m)) \\ &= \pi_{1,m} F_m(t^\star_m)/q_m - \pi_{0,m} t^B_m + \pi_{1,m}(C_m - F_m(t^\star_m)). \end{aligned} \tag{48}$$

Expression (46) is an easy consequence of (45). Finally, (48) and (45) entail

$$R_m(t^\star_m) - R_m(t^B_m) \leq \left\{ \begin{array}{ll} \pi_{1,m} C_m / q_m - \pi_{0,m} t^B_m & \text{if } t^B_m \leq t^\star_m \\ \pi_{1,m}(C_m - F_m(\Psi_m^{-1}(q_m \tau_m))) & \text{if } t^B_m \geq t^\star_m \end{array} \right. ,$$

which yields (47) because $\pi_{0,m} t^B_m = \pi_{1,m} C_m / q_m^{opt}$ by definition.

□

7.2 Proof of Theorem 3.1

Theorem 3.1 (i) follows from (47). Let us now prove (ii). First note that

$$R_m(t_m^\star) = \pi_{1,m} - \pi_{1,m} F_m(t_m^\star)(1 - q_m^{-1}). \tag{49}$$

Using (49) and the upper bound $t_m^\star = F_m(t_m^\star)(q_m \tau_m)^{-1} \leq (q_m \tau_m)^{-1}$, we obtain

$$\begin{aligned} R_m(t_m^\star) &\geq \pi_{1,m}(1 - (1 - q_m^{-1})_+ F_m(t_m^\star)) \\ &\geq \pi_{1,m}(1 - (1 - q_m^{-1})_+ F_m(q_m^{-1} \tau_m^{-1})). \end{aligned}$$

This entails (19).

7.3 Proof of Theorem 3.2

Write $\hat{t}_m$ instead of $\hat{t}_m^{FDR}$ for short. We prove the following oracle inequality, which is slightly more accurate than (21): for any $m \geq 2$,

$$\begin{aligned} R_m(\hat{t}_m) - R_m(t_m^B) \leq \pi_{1,m} \frac{\alpha_m}{1 - \alpha_m} + m^{-1} \frac{\alpha_m}{(1 - \alpha_m)^2} - \pi_{0,m} t_m^B \\ + \pi_{1,m} \gamma_m' \wedge \left\{ \gamma_m^\varepsilon + \exp\{-m\varepsilon^2 (\tau_m + 1)^{-1} (C_m - \gamma_m^\varepsilon) /4\} \right\}. \end{aligned} \tag{50}$$

Inequality (21) is a consequence of (50) where $\tau_m t_m^B$ has been lower-bounded by 0.

To establish (50), let us first write the risk of FDR thresholding as $R_m(\hat{t}_m) = T_{1,m} + T_{2,m}$, with $T_{1,m} = \pi_{0,m} \mathbb{E}(\hat{t}_m)$ and $T_{2,m} = \pi_{1,m}(1 - \mathbb{E}(F_m(\hat{t}_m)))$. In the sequel, $T_{1,m}$ and $T_{2,m}$ are examined separately.

7.3.1 Bounding $T_{1,m}$

The next result is a variation of Lemma 7.1 and Lemma 7.2 in [6].

Proposition 7.2. *The following bound holds:*

$$T_{1,m} \leq \pi_{1,m} \frac{\alpha_m}{1 - \alpha_m} + m^{-1} \frac{\alpha_m}{(1 - \alpha_m)^2}. \tag{51}$$

Proof. To prove Proposition 7.2, we follow the proof of Lemma 7.1 in [6] with slight simplifications. Remember that we have by definition $\hat{t}_m = \hat{t}_m^{BH} \vee (\alpha_m/m)$. Since α_m/m is deterministic and always smaller than the RHS in (51), and by integrating w.r.t. the label vector H, it is sufficient to prove

$$E(\hat{t}_m^{BH} | H) \leq \pi_{1,m} \frac{\alpha_m}{1 - \alpha_m} + m^{-1} \frac{\alpha_m}{(1 - \alpha_m)^2}. \tag{52}$$

Let $m_1(H) = \sum_{i=1}^m H_i$ and $m_0(H) = m - m_1(H)$. By exchangeability of $(p_i, H_i)_i$, we can assume without loss of generality that the p-values corresponding to a label $H_i = 0$ are $p_1, ..., p_{m_0(H)}$ for simplicity. Let us denote $\hat{t}_{m,0}$ the threshold of the step-up procedure applied to the p-values $p_1, ..., p_{m_0(H)}$ and using the critical values $\alpha_m(m_1(H) + k)/m$, $k = 1, ..., m_0(H)$. That is,

$$\hat{t}_{m,0} = \alpha_m(m_1(H) + \hat{k}_{m,0})/m,$$

where $\hat{k}_{m,0} = \max\{k \in \{0, 1, ..., m_0(H)\} \; : \; p_{(k)} \leq \alpha_m(m_1(H) + k)/m\}$. A classical result in multiple testing is that $\hat{t}_{m,0}$ is equal to the thresholding $\hat{t}_m^{BH}$ defined by (16), applied to the p-value family $p_i, 1 \leq i \leq m$, in which each of the p-value $p_{m_0(H)+1}, ..., p_m$ has been replaced by 0 (see, e.g., Lemma 7.1 in [25]). Moreover, since $\hat{t}_m^{BH}$ is non-increasing in each p-value, setting some p-values equal to 0 can only increase $\hat{t}_m^{BH}$. This entails

$$\mathbb{E}(\hat{t}_m^{BH} \mid H) \leq \mathbb{E}(\hat{t}_{m,0} \mid H) = \alpha_m(m_1(H) + \mathbb{E}(\hat{k}_{m,0} \mid H))/m. \tag{53}$$

Next, we may use Lemma 4.2 in [15] (by taking "$n = m_0(H)$, $\beta = \alpha_m$, $\tau = \alpha_m/m$" with their notation), to derive that for any $H \in \{0, 1\}^m$,

$$\begin{aligned}
\mathbb{E}(\hat{k}_{m,0} \mid H) &= \alpha_m \frac{m_0(H)}{m} \sum_{i=0}^{m_0(H)-1} \binom{m_0(H)-1}{i} (m_1(H) + i + 1) i! \left(\frac{\alpha_m}{m}\right)^i \\
&\leq \alpha_m \sum_{i \geq 0} (m_1(H) + i + 1)\alpha_m^i \\
&= \alpha_m (m_1(H)/(1 - \alpha_m) + 1/(1 - \alpha_m)^2).
\end{aligned}$$

The bound (52) thus follows from (53).

□

7.3.2 Bounding $T_{2,m}$

Let us consider t_m^ε the pFDR threshold associated to level $\alpha_m \pi_{0,m}(1 - \varepsilon)$. Note that by definition of t_m^ε we have $\pi_{0,m}(1-\varepsilon)G_m(t_m^\varepsilon) = t_m^\varepsilon/\alpha_m$. Here, we state the following proposition, which, combined with Proposition 7.2 establishes Theorem 3.2.

Proposition 7.3. *Let t_m^ε denote the pFDR threshold at level $\alpha_m \pi_{0,m}(1 - \varepsilon)$. Then the following bounds hold:*

$$T_{2,m} \leq \pi_{1,m}(1 - F_m(\alpha_m/m)); \tag{54}$$

$$T_{2,m} \leq \pi_{1,m}(1 - F_m(t_m^\varepsilon)) + \pi_{1,m} \exp\{-m(\tau_m + 1)^{-1}(C_m - \gamma_m^\varepsilon)\varepsilon^2/4\}. \tag{55}$$

To prove Proposition 7.3, let us first state the following lemma.

Lemma 7.4. *The following bound holds:*

$$\mathbb{P}(\hat{t}_m^{BH} < t_m^\varepsilon) \leq \exp\{-mG_m(t_m^\varepsilon)\varepsilon^2/4\}. \tag{56}$$

We can show that Lemma 7.4 implies Proposition 7.3 as follows. First, (54) is an easy consequence of $\hat{t}_m \geq \alpha_m/m$. Second, expression (55) derives from (56) because $\hat{t}_m \geq \hat{t}_m^{BH}$ and $G_m(t_m^\varepsilon) \geq \pi_{1,m}F_m(t_m^\varepsilon) \geq (\tau_m + 1)^{-1}(C_m - \gamma_m^\varepsilon)$.

Finally, we prove Lemma 7.4 by using a variation of the method described in the proof of Theorem 1 in [17] (we use Bennett's inequality instead of Hoeffding's inequality). For any $t_0 \in (0, 1)$ such that $t_0/\alpha_m - G_m(t_0) < 0$, we have

$$\begin{aligned}
\mathbb{P}(\hat{t}_m^{BH} < t_0) &\leq \mathbb{P}(\widehat{\mathbb{G}}_m(t_0) < t_0/\alpha_m) \\
&\leq \mathbb{P}(\widehat{\mathbb{G}}_m(t_0) - G_m(t_0) < t_0/\alpha_m - G_m(t_0)).
\end{aligned}$$

Next, by using Bennett's inequality (see, e.g., Proposition 2.8 in [19]) and by letting $h(u) = (1+u)\log(1+u) - u$, for any $u > 0$, we obtain

$$\mathbb{P}(\hat{t}_m^{BH} < t_0) \leq \exp\left\{ -mG_m(t_0) h\left(\frac{G_m(t_0) - t_0/\alpha_m}{G_m(t_0)} \right) \right\}.$$

Finally, for $t_0 = t_m^\varepsilon$, since we have $G_m(t_m^\varepsilon) - t_m^\varepsilon/\alpha_m = (1 - \pi_{0,m}(1-\varepsilon))G_m(t_m^\varepsilon) \geq \varepsilon G_m(t_m^\varepsilon)$, we obtain (56) by using that $h(u) \geq u^2/4$ for any $u > 0$.

8 Proofs for location and scale models

8.1 Proof of (A(F_m, τ_m))

First, assume (A'(ϕ)) and consider the location model: we easily check that

$$f_m(t) = \exp\{\phi(|\overline{D}^{-1}(t)|) - \phi(|\overline{D}^{-1}(t) - \mu_m|)\}.$$

Thus for t such that $\overline{D}^{-1}(t) > \mu_m$, we have $\log f_m(t) = \phi(\overline{D}^{-1}(t)) - \phi(\overline{D}^{-1}(t) - \mu_m) \geq \phi'(\overline{D}^{-1}(t) - \mu_m)\mu_m$, by using the convexity of ϕ. Since $\lim_{+\infty} \phi' = +\infty$, we obtain $f_m(0^+) = +\infty$. For t such that $\overline{D}^{-1}(t) < 0$, $-\log f_m(t) = \phi(-\overline{D}^{-1}(t) + \mu_m) - \phi(-\overline{D}^{-1}(t)) \geq \phi'(-\overline{D}^{-1}(t))\mu_m$. Hence we also have $f_m(1^-) = 0$. Furthermore, f_m is decreasing because ϕ is strictly convex and increasing under (A'(ϕ)).

Second, assume (A(ϕ)) and consider the scale model. In this case, we have

$$f_m(t) = \sigma_m^{-1} \exp\{\phi(\overline{D}^{-1}(t/2)) - \phi(\overline{D}^{-1}(t/2)/\sigma_m)\}.$$

Thus $f_m(1) = \sigma_m^{-1} < 1$. By using the convexity of ϕ, we have $\log(\sigma_m f_m(t)) = \phi(\overline{D}^{-1}(t/2)) - \phi(\overline{D}^{-1}(t/2)/\sigma_m) \geq (1 - \sigma_m^{-1})\overline{D}^{-1}(t/2)\phi'(\overline{D}^{-1}(t/2)/\sigma_m)$. Hence $f_m(0^+) = +\infty$. Finally, f_m is decreasing because ϕ is convex.

8.2 Proof of Proposition 4.1

Lemma 8.1. *Consider the location model with a density $d(x) = e^{-\phi(|x|)}$ for a function ϕ satisfying (A'(ϕ)). Then we have for any $m \geq 2$,*

$$\mu_m = \phi^{-1}\big(\log \tau_m + \phi(|\overline{D}^{-1}(C_m)|)\big) - \overline{D}^{-1}(C_m) \tag{57}$$

$$t_m^B \leq \tau_m^{-1} \frac{d(\overline{D}^{-1}(C_m))}{r_m^{loc}} \tag{58}$$

$$t_m^B \geq \tau_m^{-1} \frac{d(\overline{D}^{-1}(C_m))}{r_m^{loc}} \left(1 + \frac{\phi''}{\phi'^2}(\overline{D}^{-1}(C_m) + \mu_m)\right)^{-1} \quad \text{if } \phi \text{ satisfies (B}(\phi)) \tag{59}$$

$$R_m(t_m^B) \leq \pi_{1,m} \left(\frac{d(\overline{D}^{-1}(C_m))}{r_m^{loc}} + 1 - C_m \right). \tag{60}$$

If (BP) and (Sp) hold, we have $t_m^B = O\left(\pi_{1,m}/r_m^{loc}\right)$ and $R_m(t_m^B) \sim \pi_{1,m}(1 - C_m)$. If additionally (B($\phi$)) holds, we have $\tau_m t_m^B \sim \frac{d(\overline{D}^{-1}(C_m))}{r_m^{loc}}$.

Proof. First, since $C_m = F_m(t_m^B) = \overline{D}(\overline{D}^{-1}(t_m^B) - \mu_m)$, we have

$$\tau_m = f_m(t_m^B) = \exp\{\phi(|\overline{D}^{-1}(C_m) + \mu_m|) - \phi(|\overline{D}^{-1}(C_m)|)\}$$

Since $\tau_m \geq 1$ and ϕ is increasing, we get $|\overline{D}^{-1}(C_m)+\mu_m| \geq |\overline{D}^{-1}(C_m)|$. Then, we note that for any $a > 0$ and $b \in \mathbb{R}$, $|b+a| \geq |b|$ holds only if $a+b \geq 0$. This provides that $\overline{D}^{-1}(C_m)+\mu_m \geq 0$ and yields (57). Next, we have $t_m^B = F_m^{-1}(C_m) = \overline{D}(\overline{D}^{-1}(C_m) + \mu_m)$. First, using (70), we obtain that $t_m^B \leq d(\overline{D}^{-1}(C_m) + \mu_m)/\phi'(\overline{D}^{-1}(C_m) + \mu_m)$. Since $\tau_m d(\overline{D}^{-1}(C_m) + \mu_m) = d(\overline{D}^{-1}(C_m))$, we obtain (58) and then (60). Second, if ϕ satisfies ($B(\phi)$) we can apply (72) to get (59). To finish the proof, we only have to prove that ($B(\phi)$) implies that $\lim_\infty \frac{\phi''}{\phi'^2} = 0$; if ($B(\phi)$) holds then $\lim_\infty \phi'$ exists in $(0,\infty]$ and thus $h = -1/\phi'$ is non-decreasing concave with a finite limit in ∞. This entails that $h' = \phi''/(\phi')^2$ tends to zero in ∞. $\square$

Lemma 8.2. *Consider the scale model with a density $d(x) = e^{-\phi(|x|)}$ for a function ϕ satisfying* (A(ϕ)). *Then, we have for any $m \geq 2$,*

$$\log \tau_m = -\log \sigma_m + \phi(\overline{D}^{-1}(C_m/2)\sigma_m) - \phi(\overline{D}^{-1}(C_m/2)) \tag{61}$$

$$\sigma_m \geq \phi^{-1}\big(\log \tau_m + \phi(\overline{D}^{-1}(C_m/2))\big)/\overline{D}^{-1}(C_m/2) \tag{62}$$

$$t_m^B \leq \tau_m^{-1} \frac{2d(\overline{D}^{-1}(C_m/2))}{\sigma_m \phi'(\overline{D}^{-1}(C_m/2)\sigma_m)} \tag{63}$$

$$t_m^B \geq \tau_m^{-1} \frac{2d(\overline{D}^{-1}(C_m/2))}{\sigma_m \phi'(\overline{D}^{-1}(C_m/2)\sigma_m)} \left(1 + \frac{\phi''}{\phi'^2}(\overline{D}^{-1}(C_m/2)\sigma_m)\right)^{-1} \quad \textit{if } \phi \textit{ satisfies } (B(\phi)) \tag{64}$$

$$R_m(t_m^B) \leq \pi_{1,m}\left(\frac{2\overline{D}^{-1}(C_m/2)d(\overline{D}^{-1}(C_m/2))}{r_m^{sc}} + 1 - C_m\right). \tag{65}$$

In particular, if (BP) *and* (Sp) *hold, we have $\log \tau_m \sim \phi(\overline{D}^{-1}(C_m/2)\sigma_m)$, $t_m^B = O(\pi_{1,m}/r_m^{sc})$ and $R_m(t_m^B) \sim \pi_{1,m}(1-C_m)$. If additionally* ($B(\phi)$) *holds, we have $\tau_m t_m^B \sim \frac{2d(\overline{D}^{-1}(C_m/2))}{\sigma_m\phi'(\overline{D}^{-1}(C_m/2)\sigma_m)}$.*

Proof. First, since $C_m = F_m(t_m^B) = 2\overline{D}(\overline{D}^{-1}(t_m^B/2)/\sigma_m)$, we have

$$\tau_m = f_m(t_m^B) = \sigma_m^{-1} \exp\{\phi(\overline{D}^{-1}(C_m/2)\sigma_m) - \phi(\overline{D}^{-1}(C_m/2))\}$$

and thus (61) holds. Since $\log \sigma_m > 0$, we get (62). Next, using (70), we obtain that $t_m^B = F_m^{-1}(C_m) = 2\overline{D}(\overline{D}^{-1}(C_m/2)\sigma_m) \leq 2d(\overline{D}^{-1}(C_m/2)\sigma_m)/\phi'(\overline{D}^{-1}(C_m/2)\sigma_m)$. Since we have $\sigma_m\tau_m d(\overline{D}^{-1}(C_m/2)\sigma_m) = d(\overline{D}^{-1}(C_m/2))$ by (61) and by (62), we obtain (63), and then (65). Expression (64) is derived similarly by using (72). Finally, if (BP) and (Sp) holds, we obtain $\log \tau_m \sim \phi(\overline{D}^{-1}(C_m/2)\sigma_m)$ by applying (61) and by noting that $\phi(x) - \log x \sim \phi(x)$ as x tends to infinity because $\phi(x)/x \geq \phi'(1) > 0$ for $x \geq 1$. The remaining statements are then straightforward. $\square$

8.3 Proof of Corollary 4.3

Proof for (i): from Theorem 3.1 (i), to show (28), we only have to prove that $\gamma_m = (C_m - F_m(\Psi_m^{-1}(q_m\tau_m)))_+$ satisfies

$$\gamma_m \leq K_m(\log(q_m/q_m^{opt}) - \log\nu)_+/r_m. \tag{66}$$

When $q_m \leq q_m^{opt}$, this is trivial because $\gamma_m = 0$. Assume now $q_m > q_m^{opt}$ so that $\gamma_m = C_m - F_m(\Psi_m^{-1}(q_m\tau_m)) = F_m(\Psi_m^{-1}(q_m^{opt}\tau_m)) - F_m(\Psi_m^{-1}(q_m\tau_m)) \geq 0$. To prove (66), we apply Lemma 8.3 (below) with $\eta_m = C_m(1-\nu)$ to get that,

$$\begin{aligned}\log\left(\frac{\Psi_m \circ F_m^{-1}(\nu C_m))}{\Psi_m \circ F_m^{-1}(C_m)}\right) &\geq \log\nu + \frac{C_m(1-\nu)}{K_m} r_m \\ &\geq \log(q_m/q_m^{opt}),\end{aligned} \tag{67}$$

where the last inequality holds by assumption. We thus obtain $\gamma_m \leq C_m(1-\nu)$ by inverting (67) because $q_m/q_m^{opt} = \frac{\Psi_m\circ F_m^{-1}(C_m-\gamma_m)}{\Psi_m\circ F_m^{-1}(C_m)}$. We can thus apply Lemma 8.3 once again, this time for $\eta_m = \gamma_m$, we obtain

$$\log\left(\frac{q_m}{q_m^{opt}}\right) \geq \log\nu + \frac{\gamma_m}{K_m} r_m.$$

This implies (66).

Proof for (ii): We apply Theorem 3.2. Let us prove (29) for $a = 1$. Let $q_m^\varepsilon = (\alpha_m\pi_{0,m}(1-\varepsilon))^{-1} - 1 \leq (\alpha_m\pi_{0,m}(1-\varepsilon))^{-1}$ and $\gamma_m^\varepsilon = (C_m - F_m(\Psi_m^{-1}(q_m^\varepsilon\tau_m)))_+$. From the same reasoning as for (i) above, we obtain $\gamma_m^\varepsilon \leq C_m(1-\nu)$ and $\gamma_m^\varepsilon \leq K_m(\log(q_m^\varepsilon/q_m^{opt}) - \log\nu)_+/r_m$ because $r_m \geq \frac{K_m}{C_m(1-\nu)}(\log(q_m^\varepsilon/q_m^{opt}) - \log\nu)$. This yields (29) in the case $a = 1$.

Now let us prove (29) for $a = 2$. First note that $\alpha_m/m = \Psi_m^{-1}(q'_m\tau_m)$ where we let $q'_m = \tau_m^{-1} m\alpha_m^{-1} F_m(\alpha_m/m)$. Hence, $\gamma'_m = (C_m - F_m(\alpha_m/m))_+ = (C_m - F_m(\Psi_m^{-1}(q'_m\tau_m)))_+$. Assume $\alpha_m/m \leq t_m^B$ (otherwise $\gamma'_m = 0$ and the result is trivial). From the same reasoning as for (i), we can show $\gamma'_m \leq K_m(\log(q'_m/q_m^{opt}) - \log\nu)_+/r_m$. Hence the result comes from $q'_m \leq \tau_m^{-1} m\alpha_m^{-1} C_m$ because $F_m(\alpha_m/m) \leq F_m(t_m^B) = C_m$.

We now state and prove the Lemma 8.3.

Lemma 8.3. *Consider the setting of Corollary 4.3. Let η_m be such that $0 \leq \eta_m \leq C_m(1-\nu)$, for some $\nu \in (0,1)$. Then, we have*

$$\log\left(\frac{\Psi_m \circ F_m^{-1}(C_m - \eta_m)}{\Psi_m \circ F_m^{-1}(C_m)}\right) \geq \log\nu + \frac{\eta_m r_m}{K_m}. \tag{68}$$

Proof. Let us prove the location model (the scale case is similar). Let us fist note that the function $-\log\overline{D}$ is increasing on $\mathbb{R}$ and also convex on $(0,+\infty)$, because its second derivative on $(0,+\infty)$ is $d \times (-\overline{D}\phi' + d)/(\overline{D})^2$ which is non-negative by (70). Next, since $\Psi_m \circ F_m^{-1}(t) = t/\overline{D}(\overline{D}^{-1}(t) + \mu_m)$, we have

$$\begin{aligned}\log\left(\frac{\Psi_m \circ F_m^{-1}(C_m - \eta_m)}{\Psi_m \circ F_m^{-1}(C_m)}\right) &= \log\left(\frac{C_m - \eta_m}{C_m}\right) - \log\left(\frac{\overline{D}(\overline{D}^{-1}(C_m - \eta_m) + \mu_m)}{\overline{D}(\overline{D}^{-1}(C_m) + \mu_m)}\right) \\ &\geq \log\nu + (\overline{D}^{-1}(C_m - \eta_m) - \overline{D}^{-1}(C_m))\phi'(\overline{D}^{-1}(C_m) + \mu_m),\end{aligned}$$

by using that $\overline{D}^{-1}(C_m) + \mu_m > 0$ (as stated in Lemma 8.1), the convexity of $-\log \overline{D}$ on $(0, +\infty)$ and that the derivative $d/\overline{D}$ of $-\log \overline{D}$ on $(0, +\infty)$ satisfies $d/\overline{D} \geq \phi'$ (by using again(70)). Finally, since $-\overline{D}^{-1}$ is increasing and of derivative $1/d(\overline{D}^{-1}(\cdot)) \geq 1/d(0)$, we have $\overline{D}^{-1}(C_m - \eta_m) - \overline{D}^{-1}(C_m) \geq \eta_m/d(0)$. Finally note that from (57), $\phi'(\overline{D}^{-1}(C_m) + \mu_m) = r_m^{loc}$, which gives the result. □

8.4 Proof of Corollary 4.4

Let us prove (i). First note that $r_m \to \infty$ as soon as $m \to \infty$. The first claim in (i) easily derives from (28), because r_m is larger than $\frac{K_m}{C_m(1-\nu)}(\log q_m - \log \nu)$ for large m if (31) holds and because $q_m^{opt} \geq 1$. Next, we prove the second claim only in the case of the location model (the scale case in similar). Assume ($B(\phi)$) and ($C(\psi)$) for $\psi = \phi' \circ \phi^{-1}$. From above, we only have to prove that $t_m^\star$ is not asymptotically optimal whenever (31) is not fulfilled. For this, we apply Theorem 3.1 (ii) and we prove that any regime for which (31) is violated leads to (20). Up to consider a subsequence, we can assume that C_m tends to some constant $C \in (0, 1)$. It is thus sufficient to prove that $C^\star < C$ for $C^\star = \limsup_m \{(1 - q_m^{-1})_+ F_m(q_m^{-1}\tau_m^{-1})\}$.

Let us first note that the following holds from (57):

$$\begin{aligned} F_m(q_m^{-1}\tau_m^{-1}) =& \overline{D}\left(\overline{D}^{-1}(q_m^{-1}\tau_m^{-1}) - \mu_m\right) \\ =& \overline{D}\left(\overline{D}^{-1}(C_m) + \kappa_m\right), \end{aligned}$$

where $q_m = \alpha_m^{-1} - 1$ and where we let $\kappa_m = \overline{D}^{-1}(q_m^{-1}\tau_m^{-1}) - \phi^{-1}\big(\log \tau_m + \phi(|\overline{D}^{-1}(C_m)|)\big)$. Next, from ($B(\phi)$) and (74), there exists a constant $K > 0$ such that for any t small enough, $\overline{D}^{-1}(t) \geq \phi^{-1}(\log 1/t - \log \circ\phi' \circ \phi^{-1}(\log 1/t) - \log K)$. Also, from Appendix B, we can always assume that $\alpha_m \leq 1/2$, i.e. $q_m \geq 1$ for large m, and thus $q_m^{-1}\tau_m^{-1}$ necessarily converges to zero. Moreover, ϕ^{-1} is increasing and concave on $\mathbb{R}^+$, of derivative $1/\phi' \circ \phi^{-1}$. Thus we can write for m large enough,

$$\begin{aligned} \kappa_m \geq& \phi^{-1}\big(\log \tau_m + \log q_m - \log \circ\phi' \circ \phi^{-1}(\log \tau_m q_m) - \log K\big) - \phi^{-1}\big(\log \tau_m + \phi(|\overline{D}^{-1}(C_m)|)\big) \\ \geq& \frac{\log q_m - \log \circ\phi' \circ \phi^{-1}(\log \tau_m q_m) - \log K - \phi(|\overline{D}^{-1}(C_m)|}{\phi' \circ \phi^{-1}\big(\big(\log \tau_m + \log q_m - \log \circ\phi' \circ \phi^{-1}(\log \tau_m q_m) - \log K\big) \vee (\log \tau_m + \phi(|\overline{D}^{-1}(C_m)|))\big)}. \end{aligned}$$

We now use the latter bound in order to prove $C^\star < C$ in any regime for which (31) is violated.

- if α_m does not converges to 0: up to consider a subsequence, there is $\alpha_- \in (0, 1)$ such that $\alpha_m > \alpha_-$ for m large enough. Hence $\log q_m$ is bounded and we can use ($C(\psi)$)to show that $\kappa_m \sim \frac{-\log \circ\phi' \circ\phi^{-1}(\log \tau_m q_m)}{\phi' \circ \phi^{-1}\big(\log(\tau_m q_m)\big)}$ tends to zero. This implies that $C^\star \leq \limsup_m \{(1 - q_m^{-1})_+\} C \leq (1 - \alpha_-)C < C$.
- if $\alpha_m \to 0$ and $(\log q_m)/r_m^{loc}$ does not converges to zero: up to consider a subsequence, $(\log q_m)/r_m^{loc}$ converges to some $\ell \in (0, +\infty]$. First, if $\log q_m = o(\log \tau_m)$, we can use ($C(\psi)$)to show $\liminf_m \kappa_m \geq \liminf_m \frac{\log q_m}{r_m^{loc}} = \ell$. Second, if $(\log q_m)/(\log \tau_m)$ does not converges to zero, it is larger than $\delta \in (0, +\infty)$ for m large enough (up to consider a subsequence). Hence, we have $\log \tau_m \leq \delta^{-1} \log q_m$ for m large enough which entails

$\liminf_m \kappa_m \geq \liminf_m \frac{\log q_m}{\phi' \circ \phi^{-1}\left((\delta^{-1}+1)\log q_m\right)}$. Moreover, the latter is bounded away from zero because $\phi' \circ \phi^{-1}((\delta^{-1}+1)\log q_m) = O(\log q_m)$ by using ($C(\psi)$). Finally, in any case, we obtain that $\liminf_m \kappa_m > 0$ and thus $C^\star = \overline{D}\left(\overline{D}^{-1}(C) + \liminf_m \kappa_m\right) < C$.

This concludes the proof for (i).

Let us now prove (ii). First, we consider the sparsity regime where $m/\tau_m \geq (\log m)^{1+\theta}$ for some $\theta > 0$. This condition implies that for any $\kappa > 0$, $e^{-\kappa m/\tau_m}$ tends to zero faster than any power function $\tau_m^{-\lambda}$, $\lambda > 0$. In particular, since by assumption $\Psi(x) = O(e^{\lambda x})$ for $x \to +\infty$, $e^{-\kappa m/\tau_m}$ converges to zero faster than $1/r_m$. In the second sparsity regime where $m/\tau_m \to \ell \in (0, +\infty)$, we have m/τ_m which is a bounded sequence. Finally, in any of the two sparsity regimes, the result follows from Corollary 4.3 (ii), because $R_m(t_m^B) \sim \tau_m^{-1}(1-C)$ and $\pi_{1,m} \gg \alpha_m/m$.

8.5 Proof for Section 4.4

In the Laplace case, some useful relations are reported in Table 4. Also remember that from Lemma 8.2, we have

$$\log \tau_m + \log \sigma_m = (\sigma_m - 1)\log(1/C_m), \tag{69}$$

that is, $\tau_m \sigma_m = C_m^{1-\sigma_m}$. Furthermore, under (BP) and (Sp), we have $\log \tau_m \sim \log(1/C_m)\sigma_m$.

$\phi(x)$	$x + \log 2$	$F_m(t)$	$t^{\sigma_m^{-1}}$
$d(x)$	$e^{-x}/2$	$\Psi_m(t)$	$t^{\sigma_m^{-1}-1}$
$\overline{D}(x)$	$e^{-x}/2$	$\Psi_m^{-1}(v)$	$v^{1/(\sigma_m^{-1}-1)}$
$\overline{D}^{-1}(u)$	$-\log(2u)$	$F_m(\Psi_m^{-1}(v))$	$(1/v)^{1/(\sigma_m-1)}$

Table 4: Some calculations for the Laplace scale model. $x \geq 0$; $t \in (0,1)$; $v > 0$; $u \leq 1/2$.

Let us first prove Proposition 4.5 and let us start by proving (33). By definition, $R_m(t_m^\star) - R_m(t_m^B) = C_m \pi_{1,m}(Z_{1,m} + Z_{2,m})$, where $Z_{1,m} = \tau_m C_m^{-1}(\Psi_m^{-1}(q_m\tau_m)) - t_m^B)$ and $Z_{2,m} = 1 - C_m^{-1} F_m(\Psi_m^{-1}(q_m\tau_m))$. On the one hand, since $t_m^B = (C_m)^{\sigma_m}$ and using (69) twice, we get

$$\begin{aligned}
Z_{1,m} &= \tau_m C_m^{-1+\sigma_m}\left((C_m)^{-\sigma_m}\exp\left(-\frac{\log(q_m\tau_m)}{1-\sigma_m^{-1}}\right) - 1\right) \\
&= \sigma_m^{-1}\left(\exp\left(-\frac{\log q_m + \log \tau_m + (\sigma_m - 1)\log C_m}{1-\sigma_m^{-1}}\right) - 1\right) \\
&= \sigma_m^{-1}\left(\exp\left(-\frac{\log q_m - \log \sigma_m}{1-\sigma_m^{-1}}\right) - 1\right).
\end{aligned}$$

On the other hand, by using again (69), we obtain

$$\begin{aligned}
Z_{2,m} &= 1 - \exp\left(-\frac{\log q_m + \log \tau_m + (\sigma_m - 1)\log C_m}{\sigma_m - 1}\right) \\
&= 1 - \exp\left(-\frac{\log q_m - \log \sigma_m}{\sigma_m - 1}\right).
\end{aligned}$$

This implies, by denoting $\kappa_m = \log q_m - \log \sigma_m$ and by using the function g,

$$
\begin{aligned}
&(C_m \pi_{1,m})^{-1}(R_m(t_m^\star) - R_m(t_m^B)) \\
&= \sigma_m^{-1}\left(-1 + e^{-\kappa_m}\left(1 - \frac{\kappa_m}{\sigma_m - 1} + g\left(\frac{\kappa_m}{\sigma_m - 1}\right)\right)\right) + \frac{\kappa_m}{\sigma_m - 1} - g\left(\frac{\kappa_m}{\sigma_m - 1}\right).
\end{aligned}
$$

This leads to (33), because $e^{-\kappa_m} = \sigma_m / q_m$.

Next, we can prove (34) by applying Theorem 3.2. By using the above computation of $Z_{2,m}$, we have

$$
\begin{aligned}
\gamma_m^\varepsilon &\leq C_m \left(1 - \exp\left(-\frac{\log(\alpha_m^{-1}/\sigma_m) - \log(\pi_{0,m}(1-\varepsilon))}{\sigma_m - 1}\right)\right)_+ \\
&\leq C_m \frac{\big(\log(q_m/\sigma_m) - \log(\pi_{0,m}(1-\varepsilon))\big)_+}{\sigma_m - 1},
\end{aligned}
$$

because for any $u \in \mathbb{R}$, $(1 - e^{-u})_+ \leq u_+$. This gives (34) for $a = 1$. The case where $a = 2$ is similar:

$$
\begin{aligned}
(1 - C_m^{-1} F_m(\alpha/m))_+ &\leq \left(1 - \exp\left(-\frac{\log(\alpha_m^{-1} m) + (\sigma_m - 1)\log C_m}{\sigma_m - 1}\right)\right)_+ \\
&\leq \frac{\big(\log(\alpha_m^{-1}/\sigma_m) + \log(m/\tau_m)\big)_+}{\sigma_m - 1},
\end{aligned}
$$

by using (69). This finishes the proof of Proposition 4.5.

Second, let us prove Corollary 4.6 and more specifically the equivalence (35). Assume (BP) and (Sp) and that $\log(q_m/\sigma_m)$ has a limit in $\mathbb{R} \cup \{-\infty, +\infty\}$ (up to consider a subsequence). As both conditions entail that pFDR thresholding is asymptotically optimal, we can assume that $q_m \to \infty$ and $\log q_m = o(\log(\sigma_m))$ (see Corollary 4.4 (i)). Next, as g satisfies $g(x) = O(x^2)$ as $x \to 0$; $g(x) \sim x$ as $x \to +\infty$; $g(\log u) \sim 1/u$ as $u \to 0$, we easily check from (33) that the following holds:

- if $\log(q_m/\sigma_m) \to 0$, the relative excess risk tends to zero faster than $1/(\log \tau_m)$;
- if $\log(q_m/\sigma_m) \to \ell \in \mathbb{R}\backslash\{0\}$, the relative excess risk is of order $1/(\log \tau_m)$;
- if $\log(q_m/\sigma_m) \to -\infty$ or $\log(q_m/\sigma_m) \to +\infty$, the relative excess risk tends to zero slower than $1/(\log \tau_m)$;

This entails (35). Finally, let us prove (36). First note that $\sigma_m \sim (\log(1/C))^{-1}\beta \log m$. Hence, if the limit in (36) is zero, we have from (35), $\forall \beta \in \mathcal{B}$, $q_m \sim (\log(1/C))^{-1}\beta \log m$ (up to consider a subsequence). This is impossible as soon as $\mathcal{B}$ contains more than two elements (because q_m and the subsequence do not depend of β).

Acknowledgements

We would like to thank Guillaume Lecué for interesting discussions. The second author was supported by the French Agence Nationale de la Recherche (ANR grant references: ANR-09-JCJC-0027-01, ANR-PARCIMONIE, ANR-09-JCJC-0101-01) and by the French ministry of foreign and european affairs (EGIDE - PROCOPE project number 21887 NJ).

Supplementary material
Available on http://etienne.roquain.free.fr/publications.html

References

[1] F. Abramovich, Y. Benjamini, D. L. Donoho, and I. M. Johnstone. Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.*, 34(2):584–653, 2006.

[2] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57(1):289–300, 1995.

[3] Y. Benjamini, A. M. Krieger, and D. Yekutieli. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507, 2006.

[4] G. Blanchard, G. Lee, and C. Scott. Semi-supervised novelty detection. *J. Mach. Learn. Res.*, 11:2973–3009, 2010.

[5] G. Blanchard and E. Roquain. Adaptive false discovery rate control under independence and dependence. *J. Mach. Learn. Res.*, 10:2837–2871, 2009.

[6] M. Bogdan, A. Chakrabarti, F. Frommlet, and J. K. Ghosh. Asymptotic bayes-optimality under sparsity of some multiple testing procedures. *Ann. Statist.*, 39(3):1551–1579, 2011.

[7] M. Bogdan, J. K. Ghosh, and S. T. Tokdar. A comparison of the Benjamini-Hochberg procedure with some Bayesian rules for multiple testing. In *Beyond parametrics in interdisciplinary research: Festschrift in honor of Professor Pranab K. Sen*, volume 1 of *Inst. Math. Stat. Collect.*, pages 211–230. Inst. Math. Statist., Beachwood, OH, 2008.

[8] Z. Chi. On the performance of FDR control: constraints and a partial solution. *Ann. Statist.*, 35(4):1409–1431, 2007.

[9] D. Donoho and J. Jin. Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.*, 32(3):962–994, 2004.

[10] D. Donoho and J. Jin. Asymptotic minimaxity of false discovery rate thresholding for sparse exponential data. *Ann. Statist.*, 34(6):2980–3018, 2006.

[11] B. Efron. Microarrays, empirical Bayes and the two-groups model. *Statist. Sci.*, 23(1):1–22, 2008.

[12] B. Efron, R. Tibshirani, J. D. Storey, and V. Tusher. Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.*, 96(456):1151–1160, 2001.

[13] J. A. Ferreira and A. H. Zwinderman. On the Benjamini-Hochberg method. *Ann. Statist.*, 34(4):1827–1849, 2006.

[14] H. Finner, R. Dickhaus, and M. Roters. On the false discovery rate and an asymptotically optimal rejection curve. *Ann. Statist.*, 37(2):596–618, 2009.

[15] H. Finner and M. Roters. Multiple hypotheses testing and expected number of type I errors. *Ann. Statist.*, 30(1):220–238, 2002.

[16] Y. Gavrilov, Y. Benjamini, and S. K. Sarkar. An adaptive step-down procedure with proven FDR control under independence. *Ann. Statist.*, 37(2):619–629, 2009.

[17] C. Genovese and L. Wasserman. Operating characteristics and extensions of the false discovery rate procedure. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 64(3):499–517, 2002.

[18] C. Genovese and L. Wasserman. A stochastic process approach to false discovery control. *Ann. Statist.*, 32(3):1035–1061, 2004.

[19] P. Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.

[20] P. Neuvial. Intrinsic Bounds and False Discovery Rate Control in Multiple Testing Problems. hal-00460677 preprint.

[21] P. Neuvial. Asymptotic properties of false discovery rate controlling procedures under independence. *Electron. J. Stat.*, 2:1065–1110, 2008.

[22] P. Neuvial and E. Roquain. Supplement to: on false discovery rate thresholding for classification under sparsity. Submitted, 2011.

[23] E. Roquain. Type I error rate control for testing many hypotheses: a survey with proofs. *J. Soc. Fr. Stat.*, To appear.

[24] E. Roquain and M. van de Wiel. Optimal weighting for false discovery rate control. *Electron. J. Stat.*, 3:678–711, 2009.

[25] E. Roquain and F. Villers. Exact calculations for false discovery proportion with application to least favorable configurations. *Ann. Statist.*, 39(1):584–612, 2011.

[26] S. K. Sarkar. Some results on false discovery rate in stepwise multiple testing procedures. *Ann. Statist.*, 30(1):239–257, 2002.

[27] S. K. Sarkar. On methods controlling the false discovery rate. *Sankhya, Ser. A*, 70:135–168, 2008.

[28] G. R. Shorack and J. A. Wellner. *Empirical processes with applications to statistics.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[29] R. J. Simes. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754, 1986.

[30] J. D. Storey. The positive false discovery rate: a Bayesian interpretation and the q-value. *Ann. Statist.*, 31(6):2013–2035, 2003.

[31] A. C. Tamhane, W. Liu, and C. W. Dunnett. A generalized step-up-down multiple test procedure. *Canad. J. Statist.*, 26(2):353–363, 1998.

A Expressions for tails and quantiles

Lemma A.1. *Let $d(x) = e^{-\phi(|x|)}$ for any $x \in \mathbb{R}$, where ϕ is a function satisfying A(ϕ). Then $\overline{D}(x) = \int_x^{+\infty} e^{-\phi(|u|)}du$ has the following properties:*

- *for any $x > 0$, we have*

$$\overline{D}(x) \leq d(x)/\phi'(x)\,; \tag{70}$$

- *for any $t \in (0, 1/2)$ s.t. $\phi'(\overline{D}^{-1}(t)) \geq 1$, we have $-\log t > \phi(0)$ and*

$$\overline{D}^{-1}(t) \leq \phi^{-1}(-\log t)\,; \tag{71}$$

If additionally ϕ satisfies (B(ϕ)) and by letting $K = 1 + \frac{\phi''(1)}{\phi'(1)^2} > 0$, the following holds:

- *for any $x > 0$,*

$$\overline{D}(x) \geq \frac{d(x)}{\phi'(x)}\left[1 + \frac{\phi''(x)}{\phi'(x)^2}\right]^{-1}; \tag{72}$$

$$\overline{D}(x) \geq \frac{d(x)}{\phi'(x)}K^{-1} \quad \text{if } x \geq 1\,; \tag{73}$$

- *for any $t \in (0, \overline{D}(1))$ s.t. $\phi'(\overline{D}^{-1}(t)) \geq 1$, we have $-\log t > \phi(0)$ and*

$$\overline{D}^{-1}(t) \geq \phi^{-1}\bigg(\phi(0) \vee \Big\{ -\log t - \log K - \log \circ \phi' \circ \phi^{-1}(-\log t)\Big\}\bigg). \tag{74}$$

Proof. First note that $\phi'(x) > 0$ in (70) because ϕ is increasing and convex. Next, (70) holds because ϕ' is nondecreasing: $\overline{D}(x) = \int_x^{+\infty} e^{-\phi(u)}du \leq (\phi'(x))^{-1} \int_x^{+\infty} \phi'(u)e^{-\phi(u)}du = d(x)/\phi'(x)$. Expression (71) follows from (70) applied with $x = \overline{D}^{-1}(t)$. To prove (72), write for any $x > 0$,

$$\frac{\phi''(x)}{\phi'(x)^2}\overline{D}(x) \geq \int_x^{+\infty} \frac{\phi''(u)}{\phi'(u)^2}e^{-\phi(u)}du = \left[-\frac{e^{-\phi(u)}}{\phi'(u)}\right]_x^{\infty} - \overline{D}(x) = \frac{d(x)}{\phi'(x)} - \overline{D}(x),$$

by using an integration by parts. Expressions (72) and (73) follow. Finally, let us prove (74). From (73), we get $Kt\phi'(\overline{D}^{-1}(t)) \geq e^{-\phi(\overline{D}^{-1}(t))}$ and thus $-\log(Kt) - \log \circ \phi'(\overline{D}^{-1}(t)) \leq \phi(\overline{D}^{-1}(t))$. Hence, we can conclude by using (72). $\square$

B A sub-optimality result

The next proposition states a sub-optimality result when choosing a recovery parameter $q_m \leq q^+ < 1$, that is, a level $\alpha_m \geq \alpha_- > 1/2$.

Proposition B.1. *Under Assumption (A(F_m, τ_m)), let us choose $q_m \leq 1$ (i.e., $\alpha_m \geq 1/2$) in the pFDR threshold $t^\star_m$. Then we have for any $m \geq 2$,*

$$R_m(t^\star_m) \geq R_m(t^B_m)(C_m(1/q_m - 1) + 1). \tag{75}$$

In particular, under (BP)*, if* $q_m \le q_+ < 1$ *(i.e.,* $\alpha_m \ge \alpha_- > 1/2$*),*

$$\liminf_m \{R_m(t^\star_m)/R_m(t^B_m)\} > 1$$

and $t^\star_m$ *is not asymptotically optimal.*

Proof. First, since $F_m(t) = t\Psi_m(t)$,

$$\begin{aligned} R_m(t^B_m) &= \pi_{0,m} t^B_m + \pi_{0,m}\tau_m^{-1}(1 - t^B_m \Psi_m(t^B_m)) \\ &= \pi_{0,m} t^B_m (1 - \tau_m^{-1}\Psi_m(t^B_m)) + \pi_{0,m}\tau_m^{-1} \\ &\le \pi_{0,m}\tau_m^{-1}, \end{aligned} \tag{76}$$

because $\Psi_m(t^B_m) \ge f_m(t^B_m) = \tau_m$ from the concavity of F_m.

Second, assuming $q_m \le 1$, we have $\Psi_m(t^B_m) \ge \tau_m \ge q_m\tau_m = \Psi(t^\star_m)$. Hence $t^B_m \le t^\star_m$ and $F_m(t^\star_m) \ge C$. By using (49), we get $R_m(t^\star_m) \ge \pi_{0,m}\tau_m^{-1}(C(1/q_m - 1) + 1)$, which, combined with (76), leads to (75). □